

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Soichi TOYAMA

Atty. Docket No.: 107156-00206

Serial No.: New Application

Examiner: Not Assigned

Filed: September 22, 2003

Art Unit: Not Assigned

FOR: APPARATUS AND METHOD FOR SPEECH RECOGNITION

CLAIM FOR PRIORITY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313

September 22, 2003

Sir:

The benefit of the filing dates of the following prior foreign applications in the following foreign country is hereby requested for the above-identified patent application and the priority provided in 35 U.S.C. §119 is hereby claimed:

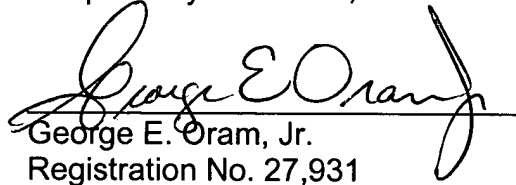
Japanese Patent Application No. 2002-271670 filed on September 18, 2002

In support of this claim, a certified copy of said original foreign application is filed herewith.

It is requested that the file of this application be marked to indicate that the requirements of 35 U.S.C. §119 have been fulfilled and that the Patent and Trademark Office kindly acknowledge receipt of these document.

Please charge any fee deficiency or credit any overpayment with respect to this paper to Deposit Account No. 01-2300.

Respectfully submitted,


George E. Oram, Jr.
Registration No. 27,931

Customer No. 004372
ARENT FOX KINTNER PLOTKIN & KAHN, PLLC
1050 Connecticut Avenue, N.W., Suite 400
Washington, D.C. 20036-5339
Tel: (202) 857-6000
Fax: (202) 638-4810

(translation)

PATENT OFFICE
JAPANESE GOVERNMENT

This is to certify that the annexed is a true copy of
the following application as filed with this office.

Date of application: September 18, 2002

Application Number: Japanese Patent Application
No. 2002-271670

[ST.10/C] : [JP2002-271670]

Applicant(s): Pioneer Corporation

Date of this certificate: June 24, 2003

Commissioner,
Japan Patent Office

Shinichiro OTA

Certificate No. 2003-3049490

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2002年 9月18日

出 願 番 号

Application Number:

特願2002-271670

[ST.10/C]:

[JP2002-271670]

出 願 人

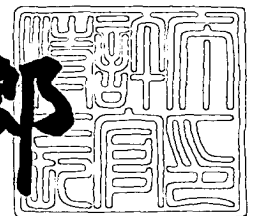
Applicant(s):

パイオニア株式会社

2003年 6月24日

特 許 庁 長 官
Commissioner,
Japan Patent Office

太田信一郎



出証番号 出証特2003-3049490

【書類名】 特許願

【整理番号】 56P0522

【提出日】 平成14年 9月18日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 3/00

【発明者】

 【住所又は居所】 埼玉県鶴ヶ島市富士見 6 丁目 1 番 1 号 パイオニア株式会社
 会社 総合研究所内

 【氏名】 外山 聡一

【特許出願人】

 【識別番号】 000005016

 【氏名又は名称】 パイオニア株式会社

【代理人】

 【識別番号】 100063565

 【弁理士】

 【氏名又は名称】 小橋 信淳

【選任した代理人】

 【識別番号】 100118898

 【弁理士】

 【氏名又は名称】 小橋 立昌

【手数料の表示】

 【予納台帳番号】 011659

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

 【物件名】 図面 1

 【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声認識装置及び音声認識方法

【特許請求の範囲】

【請求項 1】 雑音適応と話者適応とを施した合成音声モデルと、発話時の発話音声より求まる特徴ベクトル系列とを照合することによって音声認識を行う音声認識装置において、

グループ化又はクラスタリングにより、多数の音声モデルを類似性に基づいて複数のグループに分類し、当該グループ化又はクラスタリングにより同一グループに属することとなる各グループ毎の音声モデルの中から代表として求められた各グループの代表音声モデルと、前記各グループに属する音声モデルと前記代表音声モデルとの差分を前記各同一グループ毎に求めることによって得られる各グループに属する差分モデルと、前記代表音声モデルと差分モデルとを前記同一グループ毎に対応付けるグループ情報とを予め記憶する記憶手段と、

前記記憶手段に記憶されている前記同一グループ毎の代表音声モデルに対し雑音適応を施すことにより、雑音適応代表音声モデルを生成する雑音適応代表音声モデル生成手段と、

前記各グループに属している前記差分モデルと前記雑音適応代表音声モデルとを、前記グループ情報に基づいて前記同一グループ毎に合成することにより、前記同一グループ毎の合成音声モデルを生成する合成音声モデル生成手段と、

前記雑音適応を施した前記同一グループ毎の合成音声モデルに対し、発話音声より求まる特徴ベクトル系列によって話者適応を施すことにより、雑音話者適応音声モデルを生成する更新モデル生成手段と、

前記雑音話者適応音声モデルと、前記グループ情報に基づいて選択した前記雑音話者適応音声モデルの属するグループの前記雑音適応代表音声モデルとの差分から前記同一グループ毎の更新差分モデルを生成すると共に、当該生成した更新差分モデルで前記記憶手段の前記同一グループ毎の差分モデルを更新するモデル更新手段とを具備し、

音声認識に際して、前記グループ情報に基づいて選択した前記更新差分モデルの属するグループの前記代表音声モデルに雑音適応を施すことにより雑音適応代

表音声モデルを生成すると共に、当該雑音適応代表音声モデルと前記更新された更新差分モデルとを合成することで雑音適応及び話者適応を施した合成音声モデルを生成して、当該合成音声モデルと認識すべき話者音声より求まる特徴ベクトル系列とを照合することによって前記音声認識を行うことを特徴とする音声認識装置。

【請求項 2】 雑音適応と話者適応とを施した合成音声モデルと、発話時の発話音声より求まる特徴ベクトル系列とを照合することによって音声認識を行う音声認識装置において、

グループ化又はクラスタリングにより、多数の音声モデルを類似性に基づいて複数のグループに分類し、当該グループ化又はクラスタリングにより同一グループに属することとなる各グループ毎の音声モデルの中から代表として求められた各グループの代表音声モデルと、前記各グループに属する音声モデルと前記代表音声モデルとの差分を前記各同一グループ毎に求めることによって得られる各グループに属する差分モデルと、前記代表音声モデルと差分モデルとを前記同一グループ毎に対応付けるグループ情報とを予め記憶する記憶手段と、

前記記憶手段に記憶されている前記同一グループ毎の代表音声モデルに対し雑音適応を施すことにより、雑音適応代表音声モデルを生成する雑音適応代表音声モデル生成手段と、

前記各グループに属している前記差分モデルと前記雑音適応代表音声モデルとを、前記グループ情報に基づいて前記同一グループ毎に合成することにより、前記同一グループ毎の合成音声モデルを生成する合成音声モデル生成手段と、

前記合成音声モデル生成手段で生成される合成音声モデルと、認識すべき話者音声より求まる特徴ベクトル系列とを照合することにより音声認識を行う認識処理手段と、

前記同一グループ毎の合成音声モデルに対して前記話者音声より求まる特徴ベクトル系列によって話者適応を施すことにより、雑音適応と話者適応を施した雑音話者適応音声モデルを生成する更新モデル生成手段と、

前記雑音話者適応音声モデルと、前記グループ情報に基づいて選択した前記雑音話者適応音声モデルの属するグループの前記雑音適応代表音声モデルとの差分

から前記同一グループ毎の更新差分モデルを生成すると共に、当該生成した更新差分モデルで前記記憶手段の前記同一グループ毎の差分モデルを更新するモデル更新手段とを具備し、

前記認識処理手段は、音声認識が繰り返される度に前記更新モデル生成手段とモデル更新手段とによって更新される前記更新差分モデルと前記グループ情報に基づいて選択された前記更新差分モデルの属するグループの前記代表音声モデルに雑音適応を施すことで生成された雑音適応代表音声モデルとを合成することで得られる雑音適応及び話者適応が施された合成音声モデルと、認識すべき話者音声より求まる特徴ベクトル系列とを照合することにより前記音声認識を行うことを特徴とする音声認識装置。

【請求項 3】 前記モデル更新手段は、前記更新差分モデルを生成する度に、前記雑音話者適応音声モデルと前記雑音適応代表音声モデルとの類似性に基づき前記グループ情報の前記雑音話者適応音声モデルの属するグループを更に変更すると共に、前記雑音話者適応音声モデルと前記更新記憶されたグループ情報に基づいて選択された前記雑音話者適応音声モデルの属するグループの前記雑音適応代表音声モデルとの差分により、前記記憶手段の差分モデルを前記変更後のグループに即して更新することを特徴とする請求項 1 又は 2 に記載の音声認識装置。

【請求項 4】 雑音適応と話者適応とを施した合成音声モデルと、発話時の発話音声より求まる特徴ベクトル系列とを照合することによって音声認識を行う音声認識方法において、

グループ化又はクラスタリングにより、多数の音声モデルを類似性に基づいて複数のグループに分類し、当該グループ化又はクラスタリングにより同一グループに属することとなる各グループ毎の音声モデルの中から代表として求められた各グループの代表音声モデルと、前記各グループに属する音声モデルと前記代表音声モデルとの差分を前記各同一グループ毎に求めることによって得られる各グループに属する差分モデルと、前記代表音声モデルと差分モデルとを前記同一グループ毎に対応付けるグループ情報とを予め記憶手段に記憶させ、

前記記憶手段に記憶されている前記同一グループ毎の代表音声モデルに対し雑

音適応を施すことにより、雑音適応代表音声モデルを生成する雑音適応代表音声モデル生成工程と、

前記各グループに属している前記差分モデルと前記雑音適応代表音声モデルとを、前記グループ情報に基づいて前記同一グループ毎に合成することにより、前記同一グループ毎の合成音声モデルを生成する合成音声モデル生成工程と、

前記雑音適応を施した前記同一グループ毎の合成音声モデルに対し、発話音声より求まる特徴ベクトル系列によって話者適応を施すことにより、雑音話者適応音声モデルを生成する更新モデル生成工程と、

前記雑音話者適応音声モデルと、前記グループ情報に基づいて選択した前記雑音話者適応音声モデルの属するグループの前記雑音適応代表音声モデルとの差分から前記同一グループ毎の更新差分モデルを生成すると共に、当該生成した更新差分モデルで前記記憶手段の前記同一グループ毎の差分モデルを更新するモデル更新工程とを具備し、

音声認識に際して、前記グループ情報に基づいて選択した前記更新差分モデルの属するグループの前記代表音声モデルに雑音適応を施すことにより雑音適応代表音声モデルを生成すると共に、当該雑音適応代表音声モデルと前記更新された更新差分モデルとを合成することで雑音適応及び話者適応を施した合成音声モデルを生成して、当該合成音声モデルと認識すべき話者音声より求まる特徴ベクトル系列とを照合することによって前記音声認識を行うことを特徴とする音声認識方法。

【請求項 5】 雑音適応と話者適応とを施した合成音声モデルと、発話時の発話音声より求まる特徴ベクトル系列とを照合することによって音声認識を行う音声認識方法において、

グループ化又はクラスタリングにより、多数の音声モデルを類似性に基づいて複数のグループに分類し、当該グループ化又はクラスタリングにより同一グループに属することとなる各グループ毎の音声モデルの中から代表として求められた各グループの代表音声モデルと、前記各グループに属する音声モデルと前記代表音声モデルとの差分を前記各同一グループ毎に求めることによって得られる各グループに属する差分モデルと、前記代表音声モデルと差分モデルとを前記同一グ

グループ毎に対応付けるグループ情報とを予め記憶手段に記憶させ、

前記記憶手段に記憶されている前記同一グループ毎の代表音声モデルに対し雑音適応を施すことにより、雑音適応代表音声モデルを生成する雑音適応代表音声モデル生成工程と、

前記各グループに属している前記差分モデルと前記雑音適応代表音声モデルとを、前記グループ情報に基づいて前記同一グループ毎に合成することにより、前記同一グループ毎の合成音声モデルを生成する合成音声モデル生成工程と、

前記合成音声モデル生成工程で生成される合成音声モデルと、認識すべき話者音声より求まる特徴ベクトル系列とを照合することにより音声認識を行う認識処理工程と、

前記同一グループ毎の合成音声モデルに対して前記話者音声より求まる特徴ベクトル系列によって話者適応を施すことにより、雑音適応と話者適応を施した雑音話者適応音声モデルを生成する更新モデル生成工程と、

前記雑音話者適応音声モデルと、前記グループ情報に基づいて選択した前記雑音話者適応音声モデルの属するグループの前記雑音適応代表音声モデルとの差分から前記同一グループ毎の更新差分モデルを生成すると共に、当該生成した更新差分モデルで前記記憶手段の前記同一グループ毎の差分モデルを更新するモデル更新工程とを具備し、

前記認識処理工程は、音声認識が繰り返される度に前記更新モデル生成工程とモデル更新工程とによって更新される前記更新差分モデルと前記グループ情報に基づいて選択された前記更新差分モデルの属するグループの前記代表音声モデルに雑音適応を施すことで生成された雑音適応代表音声モデルとを合成することで得られる雑音適応及び話者適応が施された合成音声モデルと、認識すべき話者音声より求まる特徴ベクトル系列とを照合することにより前記音声認識を行うことを特徴とする音声認識方法。

【請求項 6】 前記モデル更新工程は、前記更新差分モデルを生成する度に、前記雑音話者適応音声モデルと前記雑音適応代表音声モデルとの類似性に基づき前記グループ情報の前記雑音話者適応音声モデルの属するグループを更に変更すると共に、前記雑音話者適応音声モデルと前記更新記憶されたグループ情報に

基づいて選択された前記雑音話者適応音声モデルの属するグループの前記雑音適応代表音声モデルとの差分により、前記記憶手段の差分モデルを前記変更後のグループに即して更新することを特徴とする請求項 4 又は 5 に記載の音声認識方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、例えば雑音適応及び話者適応等によって音声認識を行う音声認識装置及び音声認識方法に関するものである。

【0002】

【従来の技術】

音声認識の難しさの主たる原因として、一般に、音声認識すべき発話音声に背景雑音が含まれていること、及び発話者の発声気管や発話習慣などに起因する個人差があることなど上げられている。

【0003】

こうした変動要因を含んだ発話音声に基づいてロバスト（頑強）な音声認識を実現するため、HMM合成法あるいはPMC法と呼ばれる音声認識方法が研究されている（例えば、非特許文献1参照）。

【0004】

このHMM（Hidden Markov Model）合成法あるいはPMC（Parallel Model Combination）法では、実際に音声認識を行う前の前処理段階において、標準の初期音声モデル（初期音声HMM）と、背景雑音から生成した雑音モデル（発話環境雑音HMM）とを合成することによって、雑音適応を施した合成音声モデルとしての雑音適応音声モデル（雑音適応音声HMM）を生成する。

【0005】

そして、実際の音声認識に際して、発話者が発話したときの背景雑音を含んだ発話音声をケプストラム変換することによって得られる特徴ベクトル系列と、前処理段階で生成しておいた雑音適応音声モデルとを照合し、最大尤度の得られる雑音適応音声モデル等を音声認識結果として出力することとしている。

【 0 0 0 6 】

【非特許文献 1】

本間茂，高橋純一，嵯峨山茂樹、「バッチディクテーションのための教師なし話者適応」、日本音響学会講演論文集、平成 8 年 3 月、p. 5 7 - 5 8

【 0 0 0 7 】

【発明が解決しようとする課題】

ところが、従来の音声認識方法では、照合対象とすべき雑音適応音声モデル（雑音適応音声 HMM）を得るために、その全ての初期音声モデルに対して雑音適応を行う必要があり、処理量が非常に多くなるという問題があった。

【 0 0 0 8 】

また、上述したように非常に多くの処理量が必要になることから例えば初期音声モデルの種類を多くすることが困難となり、そのため、処理速度を優先する必要上、認識性能の向上を犠牲にしなければならない等の問題があった。

【 0 0 0 9 】

本発明は上記従来の問題点に鑑みてなされたものであり、例えば初期音声モデルに対して雑音適応及び話者適応等を行う際の処理量を低減し得る音声認識装置と音声認識方法を提供することを目的とする。

【 0 0 1 0 】

【課題を解決するための手段】

請求項 1 に記載の発明は、雑音適応と話者適応とを施した合成音声モデルと、発話時の発話音声より求まる特徴ベクトル系列とを照合することによって音声認識を行う音声認識装置において、グループ化又はクラスタリングにより、多数の音声モデルを類似性に基づいて複数のグループに分類し、当該グループ化又はクラスタリングにより同一グループに属することとなる各グループ毎の音声モデルの中から代表として求められた各グループの代表音声モデルと、前記各グループに属する音声モデルと前記代表音声モデルとの差分を前記各同一グループ毎に求めることによって得られる各グループに属する差分モデルと、前記代表音声モデルと差分モデルとを前記同一グループ毎に対応付けるグループ情報とを予め記憶する記憶手段と、前記記憶手段に記憶されている前記同一グループ毎の代表音声

モデルに対し雑音適応を施すことにより、雑音適応代表音声モデルを生成する雑音適応代表音声モデル生成手段と、前記各グループに属している前記差分モデルと前記雑音適応代表音声モデルとを、前記グループ情報に基づいて前記同一グループ毎に合成することにより、前記同一グループ毎の合成音声モデルを生成する合成音声モデル生成手段と、前記雑音適応を施した前記同一グループ毎の合成音声モデルに対し、発話音声より求まる特徴ベクトル系列によって話者適応を施すことにより、雑音話者適応音声モデルを生成する更新モデル生成手段と、前記雑音話者適応音声モデルと、前記グループ情報に基づいて選択した前記雑音話者適応音声モデルの属するグループの前記雑音適応代表音声モデルとの差分から前記同一グループ毎の更新差分モデルを生成すると共に、当該生成した更新差分モデルで前記記憶手段の前記同一グループ毎の差分モデルを更新するモデル更新手段とを具備し、音声認識に際して、前記グループ情報に基づいて選択した前記更新差分モデルの属するグループの前記代表音声モデルに雑音適応を施すことにより雑音適応代表音声モデルを生成すると共に、当該雑音適応代表音声モデルと前記更新された更新差分モデルとを合成することで雑音適応及び話者適応を施した合成音声モデルを生成して、当該合成音声モデルと認識すべき話者音声より求まる特徴ベクトル系列とを照合することによって前記音声認識を行うことを特徴とする。

【 0 0 1 1 】

請求項 2 に記載の発明は、雑音適応と話者適応とを施した合成音声モデルと、発話時の発話音声より求まる特徴ベクトル系列とを照合することによって音声認識を行う音声認識装置において、グループ化又はクラスタリングにより、多数の音声モデルを類似性に基づいて複数のグループに分類し、当該グループ化又はクラスタリングにより同一グループに属することとなる各グループ毎の音声モデルの中から代表として求められた各グループの代表音声モデルと、前記各グループに属する音声モデルと前記代表音声モデルとの差分を前記各同一グループ毎に求めることによって得られる各グループに属する差分モデルと、前記代表音声モデルと差分モデルとを前記同一グループ毎に対応付けるグループ情報とを予め記憶する記憶手段と、前記記憶手段に記憶されている前記同一グループ毎の代表音声

モデルに対し雑音適応を施すことにより、雑音適応代表音声モデルを生成する雑音適応代表音声モデル生成手段と、前記各グループに属している前記差分モデルと前記雑音適応代表音声モデルとを、前記グループ情報に基づいて前記同一グループ毎に合成することにより、前記同一グループ毎の合成音声モデルを生成する合成音声モデル生成手段と、前記合成音声モデル生成手段で生成される合成音声モデルと、認識すべき話者音声より求まる特徴ベクトル系列とを照合することにより音声認識を行う認識処理手段と、前記同一グループ毎の合成音声モデルに対して前記話者音声より求まる特徴ベクトル系列によって話者適応を施すことにより、雑音適応と話者適応を施した雑音話者適応音声モデルを生成する更新モデル生成手段と、前記雑音話者適応音声モデルと、前記グループ情報に基づいて選択した前記雑音話者適応音声モデルの属するグループの前記雑音適応代表音声モデルとの差分から前記同一グループ毎の更新差分モデルを生成すると共に、当該生成した更新差分モデルで前記記憶手段の前記同一グループ毎の差分モデルを更新するモデル更新手段とを具備し、前記認識処理手段は、音声認識が繰り返される度に前記更新モデル生成手段とモデル更新手段とによって更新される前記更新差分モデルと前記グループ情報に基づいて選択された前記更新差分モデルの属するグループの前記代表音声モデルに雑音適応を施すことで生成された雑音適応代表音声モデルとを合成することで得られる雑音適応及び話者適応が施された合成音声モデルと、認識すべき話者音声より求まる特徴ベクトル系列とを照合することにより前記音声認識を行うことを特徴とする。

【 0 0 1 2 】

請求項 4 に記載の発明は、雑音適応と話者適応とを施した合成音声モデルと、発話時の発話音声より求まる特徴ベクトル系列とを照合することによって音声認識を行う音声認識方法において、グループ化又はクラスタリングにより、多数の音声モデルを類似性に基づいて複数のグループに分類し、当該グループ化又はクラスタリングにより同一グループに属することとなる各グループ毎の音声モデルの中から代表として求められた各グループの代表音声モデルと、前記各グループに属する音声モデルと前記代表音声モデルとの差分を前記各同一グループ毎に求めることによって得られる各グループに属する差分モデルと、前記代表音声モデ

ルと差分モデルとを前記同一グループ毎に対応付けるグループ情報とを予め記憶手段に記憶させ、前記記憶手段に記憶されている前記同一グループ毎の代表音声モデルに対し雑音適応を施すことにより、雑音適応代表音声モデルを生成する雑音適応代表音声モデル生成工程と、前記各グループに属している前記差分モデルと前記雑音適応代表音声モデルとを、前記グループ情報に基づいて前記同一グループ毎に合成することにより、前記同一グループ毎の合成音声モデルを生成する合成音声モデル生成工程と、前記雑音適応を施した前記同一グループ毎の合成音声モデルに対し、発話音声より求まる特徴ベクトル系列によって話者適応を施すことにより、雑音話者適応音声モデルを生成する更新モデル生成工程と、前記雑音話者適応音声モデルと、前記グループ情報に基づいて選択した前記雑音話者適応音声モデルの属するグループの前記雑音適応代表音声モデルとの差分から前記同一グループ毎の更新差分モデルを生成すると共に、当該生成した更新差分モデルで前記記憶手段の前記同一グループ毎の差分モデルを更新するモデル更新工程とを具備し、音声認識に際して、前記グループ情報に基づいて選択した前記更新差分モデルの属するグループの前記代表音声モデルに雑音適応を施すことにより雑音適応代表音声モデルを生成すると共に、当該雑音適応代表音声モデルと前記更新された更新差分モデルとを合成することで雑音適応及び話者適応を施した合成音声モデルを生成して、当該合成音声モデルと認識すべき話者音声より求まる特徴ベクトル系列とを照合することによって前記音声認識を行うことを特徴とする。

【 0 0 1 3 】

請求項 5 に記載の発明は、雑音適応と話者適応とを施した合成音声モデルと、発話時の発話音声より求まる特徴ベクトル系列とを照合することによって音声認識を行う音声認識方法において、グループ化又はクラスタリングにより、多数の音声モデルを類似性に基づいて複数のグループに分類し、当該グループ化又はクラスタリングにより同一グループに属することとなる各グループ毎の音声モデルの中から代表として求められた各グループの代表音声モデルと、前記各グループに属する音声モデルと前記代表音声モデルとの差分を前記各同一グループ毎に求めることによって得られる各グループに属する差分モデルと、前記代表音声モデ

ルと差分モデルとを前記同一グループ毎に対応付けるグループ情報とを予め記憶手段に記憶させ、前記記憶手段に記憶されている前記同一グループ毎の代表音声モデルに対し雑音適応を施すことにより、雑音適応代表音声モデルを生成する雑音適応代表音声モデル生成工程と、前記各グループに属している前記差分モデルと前記雑音適応代表音声モデルとを、前記グループ情報に基づいて前記同一グループ毎に合成することにより、前記同一グループ毎の合成音声モデルを生成する合成音声モデル生成工程と、前記合成音声モデル生成工程で生成される合成音声モデルと、認識すべき話者音声より求まる特徴ベクトル系列とを照合することにより音声認識を行う認識処理工程と、前記同一グループ毎の合成音声モデルに対して前記話者音声より求まる特徴ベクトル系列によって話者適応を施すことにより、雑音適応と話者適応を施した雑音話者適応音声モデルを生成する更新モデル生成工程と、前記雑音話者適応音声モデルと、前記グループ情報に基づいて選択した前記雑音話者適応音声モデルの属するグループの前記雑音適応代表音声モデルとの差分から前記同一グループ毎の更新差分モデルを生成すると共に、当該生成した更新差分モデルで前記記憶手段の前記同一グループ毎の差分モデルを更新するモデル更新工程とを具備し、前記認識処理工程は、音声認識が繰り返される度に前記更新モデル生成工程とモデル更新工程とによって更新される前記更新差分モデルと前記グループ情報に基づいて選択された前記更新差分モデルの属するグループの前記代表音声モデルに雑音適応を施すことで生成された雑音適応代表音声モデルとを合成することで得られる雑音適応及び話者適応が施された合成音声モデルと、認識すべき話者音声より求まる特徴ベクトル系列とを照合することにより前記音声認識を行うことを特徴とする。

【 0 0 1 4 】

【発明の実施の形態】

以下、本発明の好適な実施の形態を図面を参照して説明する。

【 0 0 1 5 】

(第 1 の実施の形態)

本発明の第 1 の実施の形態を図 1 乃至図 7 を参照して説明する。なお、図 1 は、本実施形態の音声認識装置の構成を示すブロック図である。

【 0 0 1 6 】

図 1 において、本音声認識装置は、HMMを用いて音声認識を行う構成となっており、音声認識に利用する音声モデルのデータが予め記憶されている記憶部 1 と、話者環境雑音モデル生成部 2、雑音適応代表音声モデル生成部 3、合成音声モデル生成部 4、更新モデル生成部 5、モデル更新部 6、認識処理部 7 を備えて構成されている。

【 0 0 1 7 】

更に、マイクロフォン 8 で收音された收音信号 $v(t)$ を所定のフレーム期間毎にケプストラム変換し、ケプストラム領域の特徴ベクトル系列 $V(n)$ を生成して出力する音声分析部 9 と、切替スイッチ 10 が備えられている。

【 0 0 1 8 】

記憶部 1 には、予め、標準的な話者音声进行学习することによって生成された音素等のサブワード単位での多数の音声モデル等が記憶されている。

【 0 0 1 9 】

ただし、詳細については以下の説明で明らかとなるが、一般に言われている多数の初期音声モデル（標準的な話者音声を学習しただけで得られた音声モデル）がそのままの形態で記憶されているのではなく、多数の初期音声モデルの各分布（平均ベクトルと共分散行列）に対してグループ化或いはクラスタリングを施すことによって求めた、代表音声モデルの分布（C）と差分モデルの分布（D）が、代表音声モデル記憶部 1 a と差分モデル記憶部 1 b に夫々グループ分けして記憶されている。

【 0 0 2 0 】

すなわち、上述のクラスタリング等によって、多数の初期音声モデルが X 個（ $x = X$ ）のグループ $G_1 \sim G_X$ に分けられたとすると、第 1 番目（ $x = 1$ ）のグループ G_1 に属することとなった例えば q_1 個（ $q_x = q_1$ ）の初期音声モデル $S_{1,1} \sim S_{1,q_1}$ から、1 つの代表音声モデル C_1 と q_1 個の差分モデル $d_{1,1} \sim d_{1,q_1}$ を求める。

【 0 0 2 1 】

また、第 2 番目（ $x = 2$ ）のグループに属することとなった例えば q_2 個（ q

$x = q_2$) の初期音声モデル $S_{2,1} \sim S_{2,q_2}$ から、1つの代表音声モデル C_2 と q_2 個の差分モデル $d_{2,1} \sim d_{2,q_2}$ を求め、以下同様に、第 X 番目 ($x = X$) のグループに属することとなった例えば q_X 個 ($q_X = q_X$) の初期音声モデル $S_{X,1} \sim S_{X,q_X}$ から、1つの代表音声モデル C_X と、 q_X 個の差分モデル $d_{X,1} \sim d_{X,q_X}$ を求める。

【0022】

そして、図1に示すように、グループ $G_1 \sim G_X$ に属する夫々1個ずつの代表音声モデル $C_1 \sim C_X$ がグループ分けして代表音声モデル記憶部 1a に記憶され、更に、代表音声モデル C_1 に対応する q_1 個の差分モデル $d_{1,1} \sim d_{1,q_1}$ と、代表音声モデル C_2 に対応する q_2 個の差分モデル $d_{2,1} \sim d_{2,q_2}$ と、最後の代表音声モデル C_X に対応する q_X 個の差分モデル $d_{X,1} \sim d_{X,q_X}$ までの夫々の差分モデルが、各グループに対応付けられて差分モデル記憶部 1b に記憶されている。

【0023】

なお、紙面の都合上、図1には、グループ G_1 の代表音声モデル C_1 に対応する q_1 個の差分モデル $d_{1,1} \sim d_{1,q_1}$ を符号 D_1 で示し、グループ G_2 の代表音声モデル C_2 に対応する q_2 個の差分モデル $d_{2,1} \sim d_{2,q_2}$ を符号 D_2 で示し、以下同様に、グループ G_X の代表音声モデル C_X に対応する q_X 個の差分モデル $d_{X,1} \sim d_{X,q_X}$ を符号 D_X で示している。

【0024】

更に、代表音声モデル記憶部 1a に記憶された代表音声モデル $C_1, C_2 \dots C_X \dots$ と、差分モデル記憶部 1b に記憶された差分モデル $D_1, D_2 \dots D_X \dots$ とを対応付けて管理するためのグループ情報が、グループ情報記憶部 1c に記憶されている。

【0025】

図2は、既述した X 個 ($x = X$) のグループ $G_1 \sim G_X$ に対応する代表音声モデル $C_1 \sim C_X$ と、それら代表音声モデル $C_1 \sim C_X$ に対応する差分モデル $D_1 \sim D_X$ の生成原理を概念的に示した図であり、同図を参照してその生成原理を説明することとする。

【 0 0 2 6 】

まず、既述した多数の初期音声モデル（初期音声HMM）の分布 S をグループ化或いはクラスタリングすることで、類似した初期音声モデル毎にグループ分けし、更に上述のグループ情報を作成する。

【 0 0 2 7 】

ここで、グループ化の方法として、L B G 法やスプリット法などのクラスタリング手法を用い、初期音声モデルの各分布の平均ベクトルの類似性に基づいてクラスタリングを行う。

【 0 0 2 8 】

また、例えば母音モデルと子音モデルの2つのグループに分けるというように、各モデルに対応する音韻の類似性などの事前情報に基づいてグループ分けを行うようにしてもよい。

【 0 0 2 9 】

また、これら前者の手法と後者の手法を併用して、初期音声モデルをグループ分けしてもよい。

こうしてクラスタリングすると、図2に模式的に示すようなグループ分けが可能となる。

【 0 0 3 0 】

すなわち、図2において、例えば第 x 番目のグループ G_x に属することとなった音声モデルを例示列挙すれば、グループ G_x に属する第1番目の音声モデルを $S_{x, 1}$ で表すと、その平均ベクトル $\mu S_{x, 1}$ と共分散行列 $\sigma d_{x, 1}$ ($= \sigma S_{x, 1}$) から成る分布が音声モデル $S_{x, 1}$ であり、第2番目の音声モデルを $S_{x, 2}$ で表すと、その平均ベクトル $\mu S_{x, 2}$ と共分散行列 $\sigma d_{x, 2}$ ($= \sigma S_{x, 2}$) から成る分布が音声モデル $S_{x, 2}$ であり、以下同様に、第 x 番目の音声モデルを S_{x, q_x} で表すと、その平均ベクトル $\mu S_{x, q_x}$ と共分散行列 $\sigma d_{x, q_x}$ ($= \sigma S_{x, q_x}$) から成る分布が音声モデル S_{x, q_x} となる。

【 0 0 3 1 】

また、他のグループ G_1 , G_2 等に属することとなった音声モデルについても同様に、平均ベクトルと共分散行列から成る分布が音声モデルである。

【0032】

次に、各グループ $G_1 \sim G_X$ についての各代表音声モデル $C_1 \sim C_X$ の求め方を説明する。なお、説明に便宜上、図2中に示す第 x 番目のグループ G_x の代表音声モデル C_x を求める場合を代表して説明することとする。

【0033】

代表音声モデル C_x は、図2中に示す基点 Q から伸びる平均ベクトル μC_x と、その平均ベクトル μC_x に対する共分散行列 σC_x の分布（図中楕円で示す）として求める。

【0034】

したがって、代表音声モデル C_x を、 $C_x (\mu C_x, \sigma C_x)$ で表すこととすると、まず、平均ベクトル μC_x は、

【数1】

$$\mu C_x = (1/q_x) \cdot \sum_{y=1}^{q_x} \mu S_{x,y} \quad \dots (1)$$

によって求める。

【0035】

更に、共分散行列 σC_x は、

【数2】

$$\sigma S_x = (1/q_x) \cdot \sum_{y=1}^{q_x} \sigma_{x,y} + (1/q_x) \cdot \sum_{y=1}^{q_x} (\mu_{x,y} - \mu C_x) \cdot (\mu_{x,y} - \mu C_x)^T \quad \dots (2)$$

より求める。

【0036】

なお、上記式(1)(2)において、変数 x は、第 x 番目のグループ G_x であることを示し、変数 y は、グループ G_x に属する各音声モデル $S_{x,y}$ ($1 \leq y \leq q_x$)を示し、変数 q_x は、グループ G_x に属することとなった音声モデル $S_{x,y}$ の総数を示している。

【0037】

そして、他のグループ G_1, G_2 等に属する音声モデルについても上記式(1)(2)に適用し、夫々のグループの平均ベクトルと共分散行列を演算すること

により、他のグループの代表音声モデルを求める。

【0038】

次に、各グループ $G_1 \sim G_X$ に対応する差分モデル $D_1 \sim D_X$ を次式 (3) (4) に基づいて演算する。

【0039】

説明の便宜上、図 2 中に示す第 x 番目のグループ G_x の差分モデル D_x 、すなわち、 $d_{x,1}, d_{x,2} \sim d_{x,q_x}$ を求める場合を代表して述べると、

【数 3】

$$\mu_{d_{x,y}} = \mu_{S_{x,y}} - \mu_{C_x} \quad \dots (3)$$

によって、平均ベクトル $\mu_{d_{x,y}}$ を求める。更に、

【数 4】

$$\sigma_{d_{x,y}} = \sigma_{S_{x,y}} \quad \dots (4)$$

によって共分散行列 $\sigma_{d_{x,y}}$ を求める。

【0040】

なお、上記式 (3) (4) 中の変数 x は、第 x 番目のグループ G_x であることを示し、変数 y は、グループ G_x に属する各音声モデル $S_{x,y}$ ($1 \leq y \leq q_x$) を示し、変数 q_x は、グループ G_x に属することとなった音声モデル $S_{x,y}$ の総数を示している。

【0041】

そして、上記式 (3) (4) より得られた平均ベクトル $\mu_{d_{x,y}}$ と共分散行列 $\sigma_{d_{x,y}}$ を差分ベクトル $d_{x,y}$ とする。

【0042】

より具体的に述べれば、差分モデル $d_{x,1}$ は、平均ベクトル $\mu_{d_{x,1}}$ と共分散行列 $\sigma_{d_{x,1}}$ との分布、差分モデル $d_{x,2}$ は、平均ベクトル $\mu_{d_{x,2}}$ と共分散行列 $\sigma_{d_{x,2}}$ との分布、以下同様に、差分モデル $d_{x,y}$ ($y = q_x$) は、平均ベクトル $\mu_{d_{x,y}}$ と共分散行列 $\sigma_{d_{x,y}}$ との分布となり、それによって総計 q_x 個の差分モデル $d_{x,1} \sim d_{x,y}$ を求めることになる。

【0043】

こうして求めた代表音声モデル $C_1 \sim C_X$ と、差分モデル D_1 ($d_{1,1} \sim d$

$1, q_1) \sim D_X (d_{X, 1} \sim d_{X, q_X})$ が、各グループ $G_1 \sim G_X$ に対応付けられて、代表音声モデル記憶部 1a と差分モデル記憶部 1b に予め記憶されている。

【0044】

したがって、より一般的に表現すれば、図3に模式的に示すように、第 x 番目のグループ G_x に属する第 y 番目の差分モデル $d_{x, y}$ と、その差分モデル $d_{x, y}$ の属するグループ G_x の代表音声分布 C_x とを合成することにより、差分モデル $d_{x, y}$ に対応する初期音声モデル $S_{x, y}$ が求まるという関係に基づいて、各グループ G_x ($1 \leq x \leq X$) の代表音声モデル C_x ($1 \leq x \leq X$) と差分モデル D_x ($1 \leq x \leq X$) が記憶部 1a, 1b に記憶され、更にグループ情報によってグループ毎に対応付けて管理されている。

【0045】

なお、本実施形態では、平均ベクトルに対しては加算、共分散行列に対しては単なる置き換えとすることで、上述した合成を実現することとしている。すなわち、

【数5】

$$\mu_{d_{x,y}} + \mu_{C_x} = \mu_{S_{x,y}} \quad \dots (5)$$

$$\sigma_{d_{x,y}} = \sigma_{S_{x,y}} \quad \dots (6)$$

で表される関係式に従った合成処理によって、上述の合成を行うこととしている。

【0046】

なお、理解し易くするために、説明の便宜上、初期音声モデルの各分布 $S_{x, y}$ はグループ x の y 番目の分布という番号付けを行って識別したが、実際には各 HMM に対応付けられている。よって、差分モデルの各分布も同じように各 HMM に対応付けられて記憶される。

【0047】

そして、各音声 HMM に対応付けられて記憶されている初期音声モデルの各分布とその分布の属するグループとの対応関係を表すグループ情報 B もグループ情報記憶部 1c に記憶されている。

【 0 0 4 8 】

例えば、HMM番号 i の状態 j の混合 k の初期音声モデルの分布を $S^m_{i j k}$ とし、それに対応する各差分モデルを $d^m_{i j k}$ とし、更に初期音声モデルの分布 $S^m_{i j k}$ と各差分モデル $d^m_{i j k}$ の属するクラスタを β とすると、グループ情報 $B^m_{i j k}$ は分布 $S^m_{i j k}$ がどのグループに属しているかを示す情報であり、

【数 6】

$$B^m_{ijk} = \beta \dots (7)$$

となっている。

【 0 0 4 9 】

これにより、初期音声モデル及び差分モデルとその属するグループとの対応関係が、クラスタ情報 B^m によって得られるようになっている。

【 0 0 5 0 】

また、後述する雑音適応代表音声モデル生成部 3 での雑音適応手法としてヤコビ適応手法を用いており、予め作成した初期雑音モデル（便宜上 N_s とする）と上述した各グループの代表音声モデル C とを HMM 合成法により合成した初期合成音声モデルで代表音声モデル C を更新記憶する。

【 0 0 5 1 】

さらに初期雑音モデル N_s と、更新記憶された各グループの代表音声モデル C と初期雑音モデル N_s とから求めた各グループのヤコビ行列 J とを記憶し、後述する雑音適応代表音声モデル生成部 3 に供給する。

【 0 0 5 2 】

次に、発話環境雑音モデル生成部 2 は、発話環境で生じる非発話期間での背景雑音に基づいて発話環境雑音モデル（発話環境雑音 HMM） N を生成する。

【 0 0 5 3 】

すなわち、発話者が未だ発話を行っていない非発話期間に、発話環境で生じる背景雑音をマイクロフォン 8 が收音する。そして、音声分析部 9 がその收音信号 $v(t)$ から所定フレーム期間毎の背景雑音の特徴ベクトル系列 $V(n)$ を生成し、更

に切替スイッチ 10 が発話環境雑音モデル生成部 2 側に切替わることによって、その特徴ベクトル系列 $V(n)$ が背景雑音の特徴ベクトル系列 $N(n)'$ として発話環境雑音モデル生成部 2 に入力される。そして、発話環境雑音モデル生成部 2 が、特徴ベクトル系列 $N(n)'$ を学習することによって、既述した発話環境雑音モデル N を生成する。

【 0 0 5 4 】

雑音適応代表音声モデル生成部 3 は、代表音声モデル記憶部 1 a に記憶されている代表音声モデル $C_1 \sim C_X$ に対して発話環境雑音モデル N で雑音適応を施し、それによって各グループ $G_1 \sim G_X$ に対応する雑音適応代表音声モデル（雑音適応代表音声 HMM） $C_1^N \sim C_X^N$ を生成して合成音声モデル生成部 4 へ供給する。

【 0 0 5 5 】

ここで、雑音適応の手法としては、一具体例として、HMM 合成法やヤコビ適応手法等を適用して、上記代表音声モデルの分布に発話環境雑音を重畳する、いわゆる雑音適応手法を用いる。

【 0 0 5 6 】

HMM 合成法の場合は、発話環境雑音モデル N と各グループの代表音声モデル C_X とを用いて各グループの雑音適応代表音声モデル C_X^N を算出する。

【 0 0 5 7 】

ヤコビ適応手法の場合は前述のように、初期合成モデルで更新記憶されている各グループの代表音声モデル C_X と初期雑音 N_s と発話環境雑音モデル N と各グループのヤコビ行列 J とを用いて雑音適応代表音声モデル C_X^N を求める。

【 0 0 5 8 】

より一般的に、グループ G_X の代表音声モデル C_X に対し雑音適応を行う場合を述べると、背景雑音を定常と仮定し雑音モデル N を 1 状態・1 混合のモデルとした場合、上述のように HMM 合成法やヤコビ適応手法を用いた雑音適応処理により、代表音声モデル C_X は雑音適応代表音声モデル C_X^N に雑音適応され、その平均ベクトルは μC_X^N に、共分散行列は σC_X^N にそれぞれ変換される。

【 0 0 5 9 】

雑音モデル N を2状態以上あるいは2混合以上とすると、代表音声モデル C_x は2つ以上の雑音適応分布に対応することになるが、その場合、代表音声モデル C_x は、 $C_{x, 1}^N, C_{x, 2}^N, \dots$ に対応することになる。

【0060】

次に、合成音声モデル生成部4は、差分モデル記憶部1bに記憶されている各差分モデル（図中、 D で示す）と既述した各雑音適応代表音声モデル（図中、 C^N で示す）とを各グループ $G_1 \sim G_X$ に対応させて合成することにより、複数個の合成音声モデル（合成音声HMM） M を生成する。

【0061】

すなわち、一般的表現で述べると、雑音適応代表音声モデル生成部3において各グループ G_x （ $1 \leq x \leq X$ ）に対応する雑音適応代表音声モデル C_x^N （ $1 \leq x \leq X$ ）が生成されると、合成音声モデル生成部4は、グループ G_x の雑音適応代表音声モデル C_x^N （ $1 \leq x \leq X$ ）に、既述した差分モデル $d_{x, 1} \sim d_{x, y}$ （ $y = q_x$ ）を合成することにより、初期音声モデル $S_{x, 1} \sim S_{x, y}$ に対して雑音適応を施したのと等価な複数個 q_x の合成音声モデル $M_{x, 1} \sim M_{x, y}$ を生成する。

【0062】

図4は、こうして生成される複数の合成音声モデル M の構成を模式的に示した図であり、代表例として、グループ G_x に属する代表音声モデル C_x と差分モデル $d_{1, 1} \sim d_{1, y}$ （ $y = q_x$ ）から生成される合成音声モデル $M_{1, 1} \sim M_{1, y}$ の構成を示している。

【0063】

なお、図4は、理解し易くするため、上述の共分散行列を考慮せずに合成を行ったものとして簡略化して示されている。

【0064】

まず、合成音声モデル $M_{x, y}$ の平均ベクトルを $\mu M_{x, y}$ 、共分散行列を $\sigma M_{x, y}$ とする。ここで、雑音適応代表音声モデルと初期音声モデルとの合成方法として、雑音適応による代表音声モデルの分散の変動を考慮しない場合、合成音声モデル $M_{x, y}$ の平均ベクトル $\mu M_{x, y}$ と共分散行列 $\sigma M_{x, y}$ を、

【数 7】

$$\mu M_{x,y} = \mu d_{x,y} + \mu C_x^N \quad \dots (8)$$

$$\sigma M_{x,y} = \sigma d_{x,y} \quad \dots (9)$$

によって求める。また、雑音適応による代表音声モデルの共分散行列の変動も考慮する場合には、合成音声モデル $M_{x,y}$ の平均ベクトル $\mu M_{x,y}$ と共分散行列 $\sigma M_{x,y}$ を、

【数 8】

$$\mu M_{x,y} = \mu d_{x,y} + \sigma C_x^N (1/2) \cdot \sigma C_x^N (-1/2) \cdot \mu C_x^N \quad \dots (10)$$

$$\sigma M_{x,y} = \sigma C_x^N \cdot \sigma C_x^N (-1) \cdot \sigma d_{x,y} \quad \dots (11)$$

によって求める。

【0065】

ただし、音声認識性能への影響の最も大きい要因は、分布の平均ベクトル $\mu M_{x,y}$ であることから、共分散行列の分散の適応を行わない上記式 (8) (9) に基づいて、合成音声モデル $M_{x,y}$ の平均ベクトル $\mu M_{x,y}$ と共分散行列 $\sigma M_{x,y}$ を求める。本実施形態では、上記式 (8) (9) に基づいて合成音声モデル $M_{x,y}$ の平均ベクトル $\mu M_{x,y}$ と共分散行列 $\sigma M_{x,y}$ を求めており、それにより演算の処理量を低減しつつ、雑音適応性能を得ることを可能にしている。

【0066】

なお、詳細については後述するが、差分モデル記憶部 1b に記憶されている差分モデル $D_1 (d_{1,1} \sim d_{1,q_1})$, $D_2 (d_{2,1} \sim d_{2,q_2}) \dots D_x (d_{x,1} \sim d_{x,q_x}) \dots$ は、更新モデル生成部 5 とモデル更新部 6 で生成される更新差分モデルによって更新されるようになっている。

【0067】

このため説明の便宜上、図 1 中には、更新前の差分モデルを D 、更新後の差分モデルを D'' で示すと共に、更新前の差分モデル D と雑音適応代表音声モデル C^N とで合成される合成音声モデルを M とし、更新差分モデル D'' と雑音適応代表音声モデル C^N とで合成される合成音声モデルを M'' として示している。

【0068】

次に、更新モデル生成部 5 は、MLLR や MAP 法などの話者適応法によって、合成音声モデル M を特徴ベクトル系列 $V(n)$ で話者適応し、それによって雑音話者適応音声モデル（雑音話者適応音声 HMM） R を生成する。

【0069】

すなわち本実施形態では、話者適応に際して、話者適応を行うのに好適なテキスト文章等を話者に読み上げてもらう。

【0070】

更新モデル生成部 5 は、その発話期間にマイクロフォン 8 で收音され音声分析部 9 から出力される発話音声の特徴を有する所定フレーム期間毎の特徴ベクトル系列 $V(n)$ を切替スイッチ 10 を介して入力（図 1 中、点線で示す経路を通じて入力）すると共に、合成音声モデル生成部 4 で生成された合成音声モデル M を、図 1 中の点線で示す経路を通じて入力する。そして、入力した特徴ベクトル系列 $V(n)$ によって合成音声モデル M に話者適応を施すことで、雑音話者適応音声モデル R を生成する。

【0071】

図 5 は、この雑音話者適応音声モデル R の生成原理を示した模式図であり、代表例として、グループ G_x に属する代表音声モデル C_x と差分モデル D_x ($d_{x,1} \sim d_{x,y}$) との合成を上記式 (8) (9) に基づいて行い、それによって得られる合成音声モデル $M_{x,1} \sim M_{x,y}$ から雑音話者適応音声モデル $R_{x,1} \sim R_{x,y}$ を生成する場合について示している。なお、説明の便宜上、共分散行列については示されていない。

【0072】

つまり、上記式 (8) (9) の演算を行うことにより、平均ベクトル $\mu R_{x,1}$ と共分散行列 $\sigma R_{x,1}$ （図示省略）の分布から成る雑音話者適応音声モデル $R_{x,1}$ と、平均ベクトル $\mu R_{x,2}$ と共分散行列 $\sigma R_{x,2}$ （図示省略）の分布から成る雑音話者適応音声モデル $R_{x,2}$ と、以下同様に、平均ベクトル $\mu R_{x,y}$ と共分散行列 $\sigma R_{x,y}$ （図示省略）の分布から成る雑音話者適応音声モデル $R_{x,y}$ を生成する。

【0073】

そして、残余のグループ $G_1, G_2 \dots$ 等に属する雑音話者適応音声モデルについても上記式 (8) (9) に基づいて生成し、得られた全ての雑音話者適応音声モデル R をモデル更新部 6 に供給する。

【0074】

モデル更新部 6 は、更新モデル生成部 5 で生成された雑音話者適応音声モデル R と、雑音適応代表音声モデル生成部 3 で生成された雑音適応代表音声モデル C^N と、差分モデル記憶部 1 b 中の更新前の差分モデル D とを用いて、話者適応を施した更新差分モデル D'' を生成し、その更新差分モデル D'' で更新前の差分モデル D を更新する。

【0075】

グループ G_x に属する雑音話者適応音声モデル R_x と雑音適応代表音声モデル C_x^N と更新前の差分モデル D_x に対応して求められる更新差分モデル D_x'' の生成原理を代表して説明すると、更新差分モデル D_x'' 、すなわち $d_{x,1}'' \sim d_{x,y}''$ の各平均ベクトルを $\mu d_{x,1}'' \sim \mu d_{x,y}''$ 、共分散行列を $d_{x,1}'' \sim d_{x,y}''$ は、

【数 9】

$$\mu d_{x,y}'' = \alpha_{x,y} \cdot (\mu R_{x,y} - \sigma C_x^N(-1/2) \cdot \sigma C_x^N(-1/2) \cdot \mu C_x^N) + (1 - \alpha_{x,y}) \cdot \mu d_{x,y} \quad \dots (12)$$

$$\sigma d_{x,y}'' = \alpha_{x,y} \cdot (\sigma C_x \cdot \sigma C_x^N(-1) \cdot \sigma R_{x,y}) + (1 - \alpha_{x,y}) \cdot \sigma d_{x,y} \quad \dots (13)$$

によって求める。

【0076】

なお、上記式 (12) (13) は共分散行列の雑音適応を行う場合の手法を示したものであり、共分散行列の雑音適応を行わない場合には、

【数 10】

$$\mu d_{x,y}'' = \alpha_{x,y} \cdot (\mu R_{x,y} - \mu C_x^N) + (1 - \alpha_{x,y}) \cdot \mu d_{x,y} \quad \dots (14)$$

$$\sigma d_{x,y}'' = \alpha_{x,y} \cdot \sigma R_{x,y} + (1 - \alpha_{x,y}) \cdot \sigma d_{x,y} \quad \dots (15)$$

によって求める。

【0077】

また、共分散行列の話者適応も行わない場合には、

【数11】

$$\mu_{dx,y}'' = \alpha_{x,y} \cdot (\mu_{R_{x,y}} - \mu_{C_x''}) + (1 - \alpha_{x,y}) \cdot \mu_{dx,y} \quad \dots (16)$$

$$\sigma_{dx,y}'' = \sigma_{dx,y} \quad \dots (17)$$

より求める。

【0078】

話者適応では、平均ベクトルの適応効果は大きいが共分散行列の適応効果は小さい。そのため、上記式(16)(17)に示した手法により、更新差分モデル $d_{x,1}'' \sim d_{x,y}''$ の各平均ベクトル $\mu_{d_{x,1}''} \sim \mu_{d_{x,y}''}$ と共分散行列 $\sigma_{d_{x,1}''} \sim \sigma_{d_{x,y}''}$ を求めることで、演算量を低減しつつ、話者適応効果を得ることができる。このため、本実施形態では、上記式(16)(17)に基づいて更新差分モデル $d_{x,1}'' \sim d_{x,y}''$ を求めることとしている。

【0079】

尚、上記式(16)(17)中の係数 $\alpha_{x,y}$ は、雑音話者適応音声モデル $R_{x,y}$ と合成音声モデル $M_{x,y}$ から求まる更新差分ベクトル $d_{x,y}$ を適宜調整するための重み係数であり、 $0.0 \leq \alpha_{x,y} \leq 1.0$ の範囲に決められている。

【0080】

また、この重み係数 $\alpha_{x,y}$ は、予め上記範囲内の所定値に固定してもよいが、MAP推定法の重み係数のように適応が行われるたびに変更することも可能である。

【0081】

そして、図5を引用してグループ G_x に属する更新差分モデル $d_{x,1}'' \sim d_{x,y}''$ を述べると、更新差分モデル $d_{x,1}''$ は、上記式(16)中の右辺第1項から得られるベクトル $\alpha_{x,1} \cdot (\mu_{R_{x,1}} - \mu_{C_x^N})$ と第2項から得られるベクトル $(1 - \alpha_{x,1}) \cdot \mu_{d_{x,1}}$ とのベクトル和によって得られる平均ベクトル $\mu_{d_{x,1}''}$ と、上記式(17)から得られる共分散行列 $\sigma_{x,1}$

から成る分布として求まる。また、残余の更新差分モデルについても同様にして求まる。

【 0 0 8 2 】

そして、モデル更新部 6 は、全てのグループ $G_1 \sim G_X$ についての更新差分モデル $D_1'' \sim D_X''$ を求めると、記憶部 1 b に記憶されている更新前の差分モデル $D_1 \sim D_X$ を更新差分モデル $D_1'' \sim D_X''$ で更新し記憶させる。

【 0 0 8 3 】

次に、認識処理部 7 は、既述した差分モデル記憶部 1 b が更新差分モデル D'' によって更新された後、実際の音声認識が開始されるのに伴って、話者が発話した発話音声を音声認識する。

【 0 0 8 4 】

すなわち、音声認識の処理を開始すると、非発話期間内に合成音声モデル生成部 4 が、雑音適応代表音声モデル生成部 3 で生成される雑音適応代表音声モデル C^N と更新差分モデル D'' とを合成することによって、雑音適応及び話者適応を施した全グループ $G_1 \sim G_X$ の合成音声モデル M'' を生成する。

【 0 0 8 5 】

次に、話者が発話するとその発話期間に、背景雑音を含んだ話者音声の特徴ベクトル系列 $V(n)$ を音声分析部 9 が生成し切替スイッチ 10 を介して認識処理部 7 に供給する。

【 0 0 8 6 】

こうして特徴ベクトル系列 $V(n)$ が供給されると、音声認識部 7 は、特徴ベクトル系列 $V(n)$ と、合成音声モデル M'' より生成された単語や文のモデル系列とを照合し、最も高い尤度を得られる合成音声モデル M'' のモデル系列を認識結果として出力する。

【 0 0 8 7 】

次に、図 6 及び図 7 のフローチャートを参照して本音声認識装置の動作を説明する。

【 0 0 8 8 】

なお、図 6 は、音声認識を行う前に、更新差分モデル D'' を生成して差分モデル

ルDを更新する動作、図7は、更新差分モデルD”を用いて音声認識を行う際の動作を示している。

【0089】

図6において、更新処理を開始すると、まずステップS100において、雑音適応代表音声モデル生成部3が代表音声モデルCに雑音適応を施すことにより、雑音適応代表音声モデルC^Nを生成する。

【0090】

すなわち、非発話期間に収音される背景雑音の特徴ベクトル系列N(n)’が音声分析部9から発話環境雑音モデル生成部2に供給され、発話環境雑音モデル生成部2がその特徴ベクトル系列N(n)’を学習することによって発話環境雑音モデルNを生成する。

【0091】

そして、雑音適応代表音声モデル生成部3が、この発話環境雑音モデルNによって代表音声モデルCを雑音適応することにより、雑音適応代表音声モデルC^Nを生成する。

【0092】

次に、ステップS102において、合成音声モデル生成部4が、上記の雑音適応代表音声モデルC^Nと更新前の差分モデルdとを合成することにより、合成音声モデルMを生成する。

【0093】

したがって、ステップS102では、図4に示したように雑音適応の施された合成音声モデルMが生成され、未だ話者適応は施されない。

【0094】

次に、ステップS104において、更新モデル生成部5が、発話者の発した発話音声に基づいて合成音声モデルMを話者適応する。

【0095】

つまり、話者がテキスト文章等を読み上げ、その発話期間に音声分析部9から切替スイッチ10を介して発話音声の特徴ベクトル系列V(n)が更新モデル生成部5に供給されると、更新モデル生成部5がその特徴ベクトル系列V(n)によっ

て合成音声モデルMを話者適応して、雑音話者適応音声モデルRを生成する。

【0096】

したがって、ステップS104では、図5に示したように雑音適応と話者適応とが施された雑音話者適応音声モデルRが生成される。

【0097】

次に、ステップS106において、モデル更新部6が雑音話者適応音声モデルRと雑音適応代表音声モデル C^N と更新前の差分モデルDから、雑音適応と話者適応とが施された更新差分モデルD'を生成する。

【0098】

次に、ステップS108において、モデル更新部6が差分モデル記憶部1bの差分モデル（更新前の差分モデル）Dを更新差分モデルD'で更新した後、更新処理を完了する。

【0099】

このように、いわゆる初期音声モデルに対して雑音適応と話者適応を行うのではなく、代表音声モデルCについてだけ雑音適応を行い、それによって得られる雑音適応代表音声モデル C^N と差分モデルDとを合成することで合成音声モデルMを生成して話者適応を施すので、雑音適応と話者適応に要する処理量を大幅に削減することができる。

【0100】

更に、この更新処理の際、雑音適応と話者適応を施した更新差分モデルD'を生成して差分モデル記憶部1bの内容を更新しておくので、次に述べる音声認識の際の処理量も大幅に低減することができ、迅速な音声認識を可能にする。

【0101】

次に、図7を参照して音声認識の際の動作を説明する。

【0102】

同図において話者からの指示を受けると音声認識の処理を開始し、ステップS200において、雑音適応代表音声モデル生成部3が記憶部1a中の代表音声モデルCを雑音適応することにより、雑音適応代表音声モデル C^N を生成する。

【0103】

つまり、未だ話者が発話していない非発話期間内に、音声分析部 9 から出力される背景雑音の特徴ベクトル系列 $N(n)'$ を発話環境雑音モデル生成部 2 が学習して発話環境雑音モデル N を生成すると、雑音適応代表音声モデル生成部 3 がその発話環境雑音モデル N によって代表音声モデル C を雑音適応し、雑音適応代表音声モデル C^N を生成する。

【 0 1 0 4 】

次に、ステップ S 2 0 2 において、合成音声モデル生成部 4 が雑音適応代表者モデル C^N と更新差分モデル D'' とを合成し、雑音適応と話者適応とが施された合成音声モデル M'' を生成する。

【 0 1 0 5 】

次に、ステップ S 2 0 4 において、認識処理部 7 が話者音声の特徴ベクトル系列 $V(n)$ と合成音声モデル M'' から生成した単語や文のモデルとを照合して音声認識する。

【 0 1 0 6 】

つまり、話者が発話を開始すると、切替スイッチ 1 0 が認識処理部 7 側に切り替わり、その発話期間において音声分析部 9 から出力される背景雑音の重畳した発話音声の特徴ベクトル系列 $V(n)$ が認識処理部 7 に供給される。

【 0 1 0 7 】

そして、認識処理部 7 がこの特徴ベクトル系列 $V(n)$ と合成音声モデル M'' から生成した単語や文のモデルとを照合し、ステップ S 2 0 6 において最大尤度の得られる合成音声モデル M'' のモデル系列（上記単語や文に対応するモデル系列）を音声認識結果として出力する。

【 0 1 0 8 】

このように、音声認識の際にも、いわゆる初期音声モデルに対して雑音適応と話者適応を行うのではなく、雑音適応代表音声モデル C^N と更新差分モデル D'' とを合成することで雑音適応と話者適応の施された合成音声モデル M'' を生成するので、雑音適応と話者適応に要する処理量を大幅に削減することができる。

【 0 1 0 9 】

更に、従来の音声認識では、話者適応を行うこととするとその話者の発話環境

の影響によって環境適応も行われてしまうことから、その話者適応と共に環境適応がなされた音響モデルを照合対象として、発話音声の特徴ベクトル系列 $V(n)$ との照合を行うことになり、音声認識率の向上を阻害する要因となっていた。

【 0 1 1 0 】

しかし、本実施形態によれば、話者適応後の音響モデルを差分モデル化すなわち更新差分モデル D'' として生成し、その更新差分モデル D'' から照合対象としての合成音声モデル M'' を生成するので、環境適応の影響を低減することができる。これにより、雑音適応と話者適応の相乗効果が得られ、より高い音声認識率を実現することができる。

【 0 1 1 1 】

(第 2 の実施の形態)

次に、本発明の第 2 の実施形態を図 8 及び図 9 を参照して説明する。

尚、図 8 は本実施形態の音声認識装置の構成を示す図であり、図 1 と同一又は相当する部分を同一符号で示している。

【 0 1 1 2 】

図 8 において、本音声認識装置と第 1 の実施形態の音声認識装置との差異を述べると、第 1 の実施形態の音声認識装置では、図 6 及び図 7 のフローチャートを参照して説明したように、雑音適応と話者適応とを施した更新差分モデル D'' を生成した後、音声認識を行うのに対し、本実施形態の音声認識装置は、音声認識中に更新モデル生成部 5 とモデル更新部 6 が更新処理を行うことで、音声認識中に差分モデル D'' の生成を同時に行うようになっている。

【 0 1 1 3 】

次に、図 9 のフローチャートに基づいて本音声認識装置の動作を説明する。

【 0 1 1 4 】

図 9 において音声認識処理を開始すると、まずステップ S 3 0 0 において、雑音適応代表音声モデル生成部 3 が代表音声モデル C に雑音適応を施すことにより、雑音適応代表音声モデル C^N を生成する。

【 0 1 1 5 】

すなわち、話者が未だ発話を開始する前の非発話期間に収音される背景雑音の

特徴ベクトル系列 $N(n)'$ が音声分析部9から発話環境雑音モデル生成部2に供給され、発話環境雑音モデル生成部2がその特徴ベクトル系列 $N(n)'$ を学習することによって発話環境雑音モデル N を生成する。

【0116】

そして、雑音適応代表音声モデル生成部3が、この発話環境雑音モデル N によって代表音声モデル C を雑音適応することにより、雑音適応代表音声モデル C^N を生成する。

【0117】

次に、ステップS302において、合成音声モデル生成部4が、上記の雑音適応代表音声モデル C^N と更新前の差分モデル D とを合成することにより、合成音声モデル M を生成する。

【0118】

次に、ステップS304において、認識処理部7が話者音声の特徴ベクトル系列 $V(n)$ と、合成音声モデル M より生成された単語や文のモデル系列とを照合して音声認識する。

【0119】

つまり、話者が発話を開始すると、切替スイッチ10が認識処理部7側に切り替わり、その発話期間において音声分析部9から出力される発話音声の特徴ベクトル系列 $V(n)$ が認識処理部7に供給される。そして、認識処理部7がこの特徴ベクトル系列 $V(n)$ と合成音声モデル M より生成したモデル系列とを照合し、ステップS306において最大尤度の得られる合成音声モデル M のモデル系列を音声認識結果 RCG として出力する。

【0120】

更にステップS306では、上位候補の尤度情報も同時に出力し、更にその上位候補の尤度値から認識結果の確からしさ（「信頼度」という）を所定の基準に照らして決定する。

【0121】

次に、ステップS308では、上述した信頼度に基づいて、認識結果を正しいと判断し得るか否かを判断し、正しい（正解）と判断すると、ステップS310に

移行し、正解でない（正解とし得ない）と判断すると、認識終了とする。なお、既述した認識結果の信頼度の計算方法としては、様々な方法があるがここでは省略することとする。

【0122】

次に、ステップS310、S312において、更新モデル生成部5が、既述の合成音声モデルMと発話音声の特徴ベクトル系列 $V(n)$ 及び音声認識結果RCGを用いて、話者適応を行い、更にモデル更新部6が、更新差分モデル D'' を生成して更新前の差分モデルDを更新する。

【0123】

すなわち、まずステップS310において、更新モデル生成部5が、認識されたモデル系列を音声認識結果RCGによって判別し、特徴ベクトル系列 $V(n)$ によって合成音声モデルMを話者適応する。

【0124】

これにより、例えば発話者が「東京」と発話し、その単語「東京」の音声認識結果RCGが認識処理部7から出力されると、単語「東京」の合成音声モデルMに対して発話音声「東京」の特徴ベクトル系列 $V(n)$ によって話者適応が行われ、雑音適応と話者適応が施された雑音話者適応音声モデルRが生成される。

【0125】

更に、モデル更新部6が雑音話者適応音声モデルRと雑音適応代表音声モデル C^N と更新前の差分モデルDから、音声認識結果RCGに対応する更新差分モデル D'' を生成する。

【0126】

そしてステップS312において、モデル更新部6が音声認識結果RCGに対応する差分モデル（更新前の差分モデル）Dを更新差分モデル D'' で更新する。

【0127】

これにより、前述した音声認識結果RCGが単語「東京」の場合には、その「東京」という単語の更新前の差分モデルDが更新差分モデル D'' で更新される。

【0128】

このように、本実施形態の音声認識装置によれば、代表音声モデル記憶部1a

と差分モデル記憶部 1 b に予め設定されている代表音声モデル C と差分モデル D を用いて音声認識を行い、それと同時に雑音適応と話者適応を施した更新差分モデル D'' を生成することができる。

【 0 1 2 9 】

更に注目すべきは、最初の音声認識を終了した後、例えば別の日時などに音声認識を行うと、音声認識を繰り返す回数が増えるたびに、更新前の差分モデル D が次第に話者適応された更新差分モデル D'' に更新されていく。このため、図 9 中のステップ S 3 0 2 で生成される合成音声モデル M は、雑音適応と話者適応の施された合成音声モデルとなっていく。

【 0 1 3 0 】

したがって、音声認識処理部 7 は、合成音声モデル M'' と発話音声の特徴ベクトル系列 $V(n)$ とを照合して音声認識を行うようになるため、本音声認識装置の使用回数が増えるのに伴って認識率が向上するという優れた効果が得られる。

【 0 1 3 1 】

なお、第 1、第 2 の実施形態において、予め設定されている差分モデル D を更新差分モデル D'' に更新する度に、グループ情報の更新を行っても良い。

【 0 1 3 2 】

すなわち、第 1 の実施形態においてモデル更新部 6 が図 6 に示したステップ S 1 0 8 の処理を完了した後、差分モデル記憶部 1 b に記憶されることとなった更新差分モデル D'' と代表音声モデル C とを合成した合成モデル S'' と、代表音声モデル C との類似性に基づき、最も類似した代表音声モデルの属するグループのメンバーとなるようグループ情報と更新差分モデルの変更を行う。

【 0 1 3 3 】

前述のように、実際には更新差分モデル $d_{x, y}$ は、HMM 番号 i 、状態番号 j 、混合番号 k に対して $d_{i j k}^m$ の形で記憶される。

【 0 1 3 4 】

また、前述の如く $d_{i j k}^m$ の属するクラスタは、クラスタ情報 $B_{i j k}^m$ として記憶される。例えば、 $d_{i j k}^m$ の属するクラスタが β だったとする。すなわち、 $B_{i j k}^m = \beta$ とすると、 $d_{i j k}^m$ の属するクラスタの代表モデ

ルは C_{β} である。よって、HMM番号 i 、状態番号 j 、混合番号 k の合成モデル S^m_{ijk} は、 d^m_{ijk} と C_{β} を合成することにより求まる。

【0135】

ここで、 S^m_{ijk} と全ての代表音声モデルとの類似性に基づいて比較した結果、もっとも類似した音声モデルが C_{β} でなく、 C_{γ} だったとする。その場合、更新差分モデルは、

$$d^m_{ijk} = S^m_{ijk} - C_{\gamma}$$

と置き換える。またクラスタ情報も、

$$B^m_{ijk} = \gamma$$

と置き換える。

【0136】

そして、更新された差分情報・グループ情報は記憶部1cに更新記憶させる。

【0137】

なお、合成モデル S ”に対しグループ化またはクラスタリングを施し、グループ情報 B 、代表音声モデル C 、更新差分モデル D ”を更新するようにすることもできるが、クラスタリングは多くの演算量を必要とする処理であり効果的ではない。

【0138】

また、雑音適応手法としてヤコビ適応を使用する場合は代表音声モデル C の更新を行うと初期合成モデル作成などに多くの演算が必要になる。

【0139】

少ない演算量で効果を得るには、上記のように差分モデルとグループ情報のみを書き換える方法が効果的である。

【0140】

また、第2の実施形態では、図9に示したステップS310の処理を完了した後、差分モデル記憶部1bに記憶されることとなった更新差分モデル D ”と代表音声モデル C とを合成した合成モデル S ”と、代表音声モデル C との類似性に基づき、最も類似した代表音声モデルの属するグループのメンバーとなるようグループ情報と更新差分モデルの変更を行う。

【0141】

前述のように、実際には更新差分モデル $d_{x,y}^m$ は、HMM番号 i 、状態番号 j 、混合番号 k に対して d_{ijk}^m の形で記憶される。

【0142】

また、前述の如く d_{ijk}^m の属するクラスタは、クラスタ情報 B_{ijk}^m として記憶される。例えば、 d_{ijk}^m の属するクラスタが β だったとする。すなわち、 $B_{ijk}^m = \beta$ とすると、 d_{ijk}^m の属するクラスタの代表モデルは C_β である。よって、HMM番号 i 、状態番号 j 、混合番号 k の合成モデル S_{ijk}^m は、 d_{ijk}^m と C_β を合成することにより求まる。

【0143】

ここで、 S_{ijk}^m と全ての代表音声モデルとの類似性に基づいて比較した結果、もっとも類似した音声モデルが C_β でなく、 C_γ だったとする。その場合、更新差分モデルは、

$$d_{ijk}^m = S_{ijk}^m - C_\gamma$$

と置き換える。またクラスタ情報も、

$$B_{ijk}^m = \gamma$$

と置き換える。

【0144】

そして、更新された差分情報・グループ情報は記憶部 1c に更新記憶させる。

【0145】

なお、合成モデル S に対しグループ化またはクラスタリングを施し、グループ情報 B 、代表音声モデル C 、更新差分モデル D を更新するようにすることもできるが、クラスタリングは多くの演算量を必要とする処理であり効果的ではない。

【0146】

また、雑音適応手法としてヤコビ適応を使用する場合は代表音声モデル C の更新を行うと初期合成モデル作成などさらに多くの演算が必要になる。

【0147】

少ない演算量で効果を得るには、上記のように差分モデルとグループ情報のみ

を書き換える方法が効果的である。

【 0 1 4 8 】

以上に述べたように第 1，第 2 の実施形態によれば、認識処理の処理量を低減しつつ、音声認識率の更なる向上を実現することができる。

【 0 1 4 9 】

つまり、第 1 の実施形態に係る音声認識装置とその音声認識方法によれば、音声認識を行う前に、更新差分モデルを生成して記憶部 1 に記憶しておき、その更新差分モデルを利用して音声認識の処理を行う。すなわち、多数の音声モデルを類似性に基づきグループ化又はクラスタリングし、それによって得られるグループ情報とグループの代表音声モデルと差分モデルとを同一グループ毎に対応付けて記憶部 1 に記憶させておく。

【 0 1 5 0 】

そして、音声認識を行う前に、雑音適応と話者適応を施した更新差分モデルを生成し、その更新差分モデルで記憶部 1 の差分モデルを更新する。

【 0 1 5 1 】

ここで、上述の更新差分モデルで記憶部 1 の差分モデルを更新する際には、まず、記憶部 1 に記憶されている同一グループ毎の代表音声モデルに対し雑音適応を施すことにより、同一グループ毎の雑音適応代表音声モデルを生成する。

【 0 1 5 2 】

更に、雑音適応代表音声モデルと記憶部 1 の差分モデルとを、同一グループ毎に合成することによって、雑音適応を施した合成音声モデルを生成する。

【 0 1 5 3 】

更に、その雑音適応を施した合成音声モデルに対し、発話音声により求まる特徴ベクトル系列によって話者適応を施すことにより、雑音話者適応音声モデルを生成する。

【 0 1 5 4 】

そして、雑音話者適応音声モデルと雑音適応代表音声モデルとの差分から更新差分モデルを生成し、その更新差分モデルで記憶部 1 の差分モデルを更新する。

【 0 1 5 5 】

次に、音声認識の際には、記憶部 1 に記憶されている代表音声モデルを雑音適応し、それによって得られる雑音適応代表音声モデルと更新された更新差分モデルとを合成することで得られる雑音適応及び話者適応を施した合成音声モデルと、認識すべき話者の発話音声より求まる特徴ベクトル系列とを照合することにより、音声認識を行う。

【0156】

このように、代表音声モデルと差分モデルを使用して、差分モデルに対して雑音適応と話者適応を施すことによって更新差分モデルを生成しておき、音声認識の際、代表音声モデルを雑音適応した雑音適応代表音声モデルと更新差分モデルとを合成することにより、話者の発話音声より求まる特徴ベクトル系列との照合を行うための合成音声モデルを少ない処理量で生成することを可能にする。

【0157】

つまり、雑音適用と話者適応の処理を、音声認識を行うために必要となる多数の音声モデルの全てについて行うのではなく、多数の音声モデルを代表音声モデルと差分モデルとにいわゆる分解しておいて、これら代表音声モデルと差分モデルに対し、雑音適用と話者適応を施して合成等することで、話者の発話音声より求まる特徴ベクトル系列との照合を行うための合成音声モデルを生成する。これにより、処理量の大幅な低減を実現する。

【0158】

また、この第 1 の実施形態では、上述の雑音話者適応モデルを求めた後、更に雑音話者適応モデルの属するグループを各雑音適応代表音声モデルとの類似性に基づき変更し、その変更が反映されるようグループ情報を更新記憶し、さらに雑音話者適応モデルと変更されたグループの雑音適応代表音声モデルとの差分を更新差分モデルとする。そして、更新された差分モデルと更新されたグループ情報に基づく代表音声モデルに雑音適応を施した雑音適応代表音声モデルとを合成することによって得られる合成音声モデルを用いて音声認識を行う。このグループ情報及び差分モデルを更新することによって、音声認識率の向上等を実現することが可能となっている。

【0159】

第 2 の実施形態に係る音声認識装置とその音声認識方法によれば、多数の音声モデルを類似性に基づきグループ化又はクラスタリングし、それによって得られるグループ情報と各グループの代表音声モデルと差分モデルとを同一グループ毎に対応付けて記憶部 1 に記憶させておく。音声認識を行う度に、その音声認識の処理中に、雑音適応と話者適応を施した更新差分モデルを生成し、その更新差分モデルで記憶部 1 の差分モデルを同一グループ毎に更新する。

【 0 1 6 0 】

そして、音声認識が繰り返される度に、更新差分モデルによる更新が行われ、次第に話者適応の効果の高くなっていく更新差分モデルと、代表音声モデルに雑音適応を施した雑音適応代表音声モデルとを合成し、その合成により得られる合成音声モデルと、話者の発話音声より求まる特徴ベクトル系列とを照合することで、音声認識を行う。

【 0 1 6 1 】

ここで、更新差分モデルで記憶部 1 の差分モデルを更新する際には、記憶部 1 に記憶されている代表音声モデルに対し雑音適応を施すことにより、雑音適応代表音声モデルを生成する。

【 0 1 6 2 】

更に、雑音適応代表音声モデルと記憶部 1 の差分モデルとを合成することによって合成音声モデルを生成する。

【 0 1 6 3 】

更に、合成音声モデルに対して話者の発話音声より求まる特徴ベクトル系列によって話者適応を施すことにより、雑音適応と話者適応を施した雑音話者適応音声モデルを生成する。

【 0 1 6 4 】

そして、雑音話者適応音声モデルと雑音適応代表音声モデルとの差分から更新差分モデルを生成し、その更新差分モデルで記憶部 1 の差分モデルを更新する。

【 0 1 6 5 】

更にまた、音声認識が繰り返される度に、最新の更新差分モデルで記憶部 1 中の古い更新差分モデルを更新していく。

【0166】

そして、更新差分モデルと、記憶部1に記憶されている代表音声モデルに雑音適応を施した雑音適応代表音声モデルとを合成し、その合成で得られる雑音適応と話者適応が施された合成音声モデル、及び、認識すべき話者の発話音声より求まる特徴ベクトル系列とを照合することにより音声認識を行う。

【0167】

このように、代表音声モデルと差分モデルを使用して、差分モデルに対して雑音適応と話者適応を施すことによって更新差分モデルを生成し、音声認識を行う度に、代表音声モデルに雑音適応を施した雑音適応代表音声モデルと更新差分モデルとを合成することにより、話者の発話音声より求まる特徴ベクトル系列との照合を行うための合成音声モデルを少ない処理量で生成することを可能にする。

【0168】

また、第2の実施形態においても、上述の雑音話者適応モデルを求めた後、更に雑音話者適応モデルの属するグループを各雑音適応代表音声モデルとの類似性に基づき変更し、その変更が反映されるようグループ情報を更新記憶し、さらに雑音話者適応モデルと変更されたグループの雑音適応代表音声モデルとの差分を更新差分モデルとする。そして、更新された差分モデルと更新されたグループ情報に基づく代表音声モデルに雑音適応を施した雑音適応代表音声モデルとを合成することによって得られる合成音声モデルを用いて音声認識を行う。このグループ情報及び差分モデルを更新することによって、音声認識率の向上等を可能にしている。

【0169】

このように、第1、第2の実施形態によれば、雑音適応代表音声モデルと差分モデルと適応発話音声を使用して、差分モデルに対して話者適応を施すことによって更新差分モデルを生成し、音声認識に際して、雑音適応代表音声モデルと更新差分モデルとの合成によって話者音声の特徴ベクトルと照合するための雑音適応及び話者適応を施した合成音声モデルを生成するようにしたので、その合成音声モデルを生成するための処理量を大幅に低減することができると共に、音声認識処理の高速化及び認識精度の向上を図ることができる。

【図面の簡単な説明】

【図 1】

第 1 の実施形態の音声認識装置の構成を示す図である。

【図 2】

代表音声モデルと差分モデルの生成原理を示す図である。

【図 3】

代表音声モデルと差分モデルと初期音声モデルの関係を示す図である。

【図 4】

雑音適応された合成音声モデルの生成原理を示す図である。

【図 5】

雑音適応と話者適応された雑音話者適応音声モデルの生成原理及び更新差分モデルの生成原理を示す図である。

【図 6】

差分モデルを更新差分モデルで更新するまでの動作を示すフローチャートである。

【図 7】

音声認識の際の動作を示すフローチャートである。

【図 8】

第 2 の実施形態の音声認識装置の構成を示す図である。

【図 9】

第 2 の実施形態の音声認識装置の動作を示すフローチャートである。

【符号の説明】

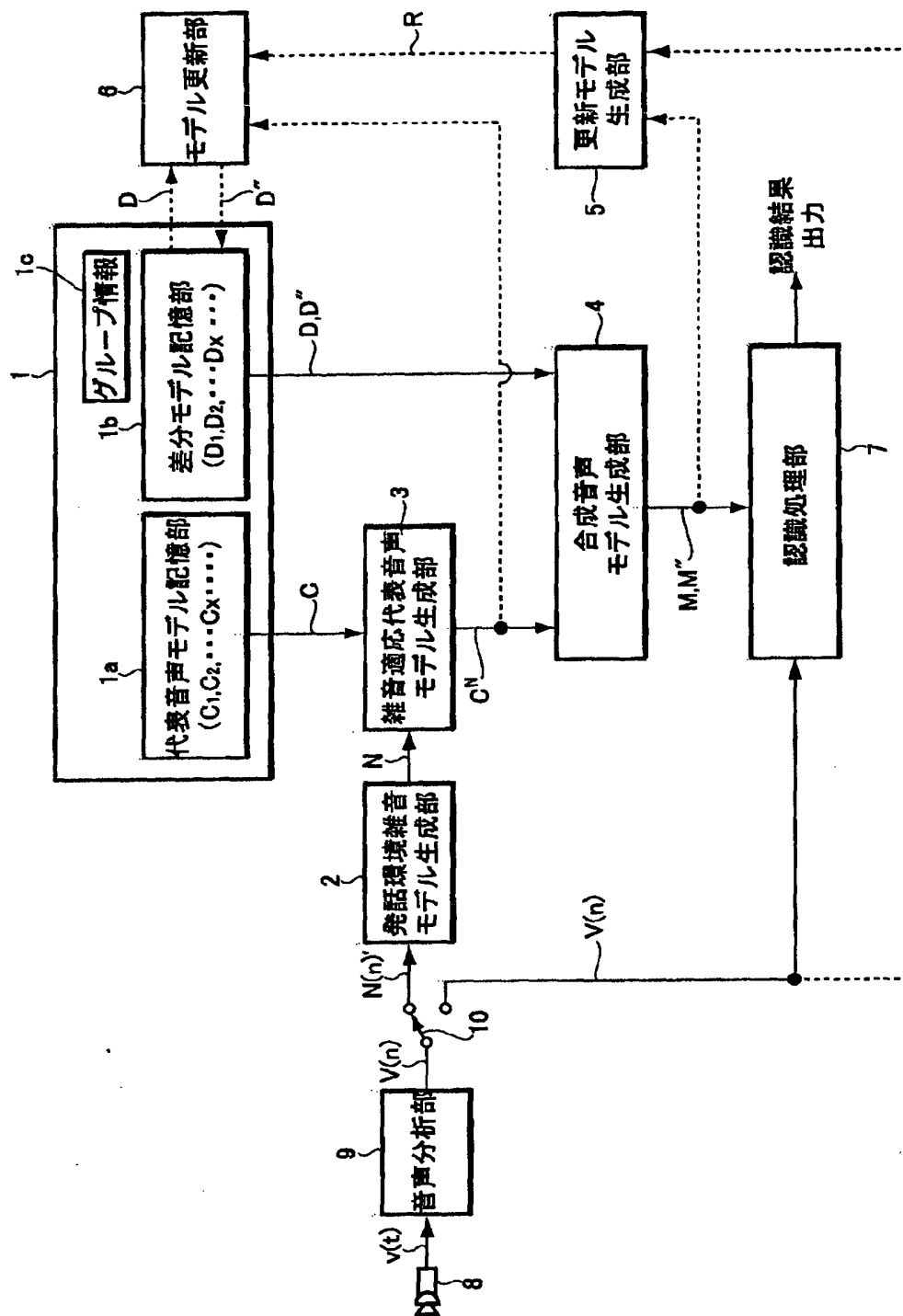
- 1 … 記憶部
- 1 a … 代表音声モデル記憶部
- 1 b … 差分モデル記憶部
- 1 c … グループ情報記憶部
- 2 … 発話環境雑音モデル生成部
- 3 … 雑音適応代表音声モデル生成部
- 4 … 合成音声モデル生成部

- 5 …更新モデル生成部
- 6 …モデル更新部
- 7 …認識処理部
- 8 …マイクロフォン
- 9 …音声分析部
- 1 0 …切替スイッチ

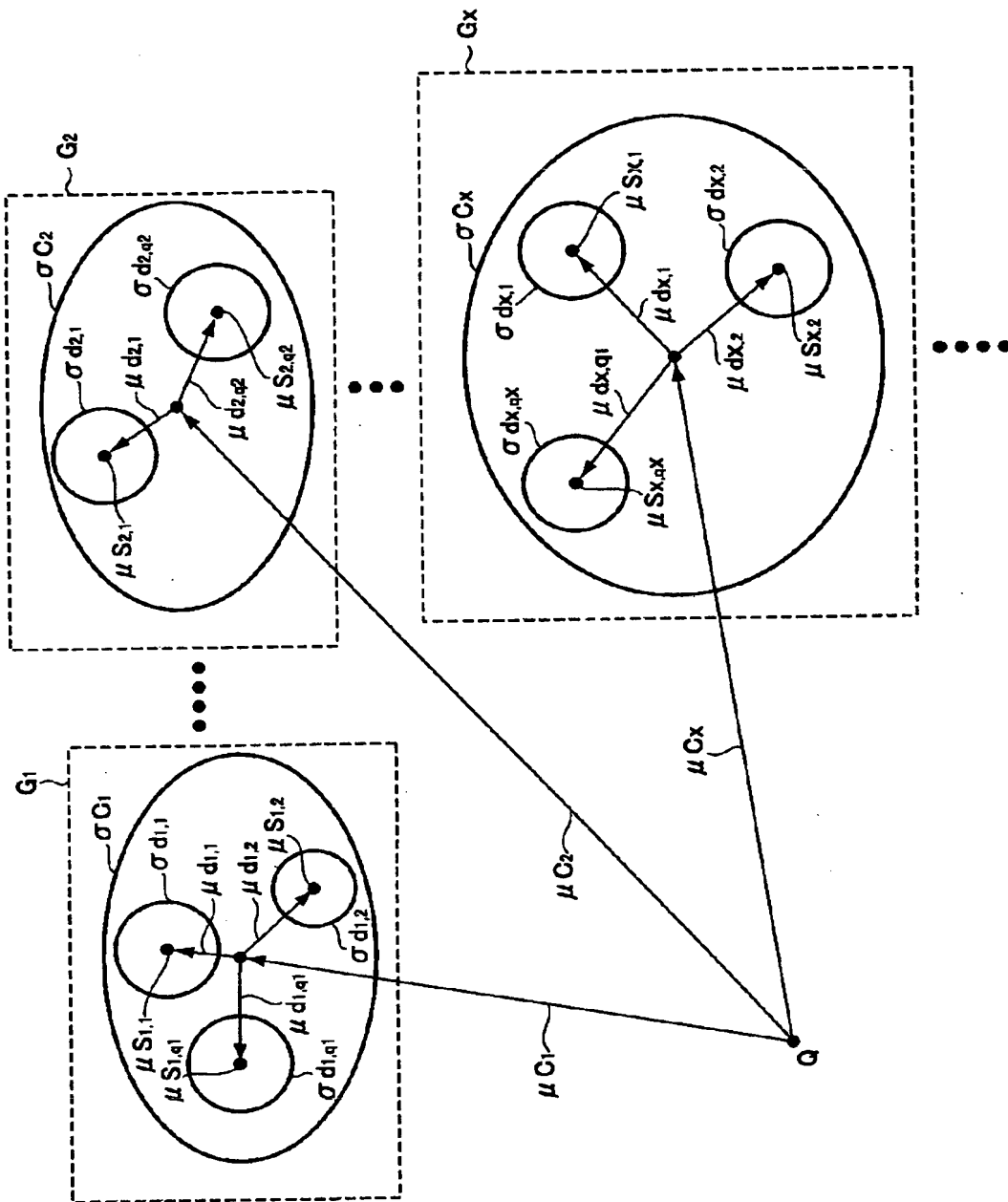
【書類名】

図面

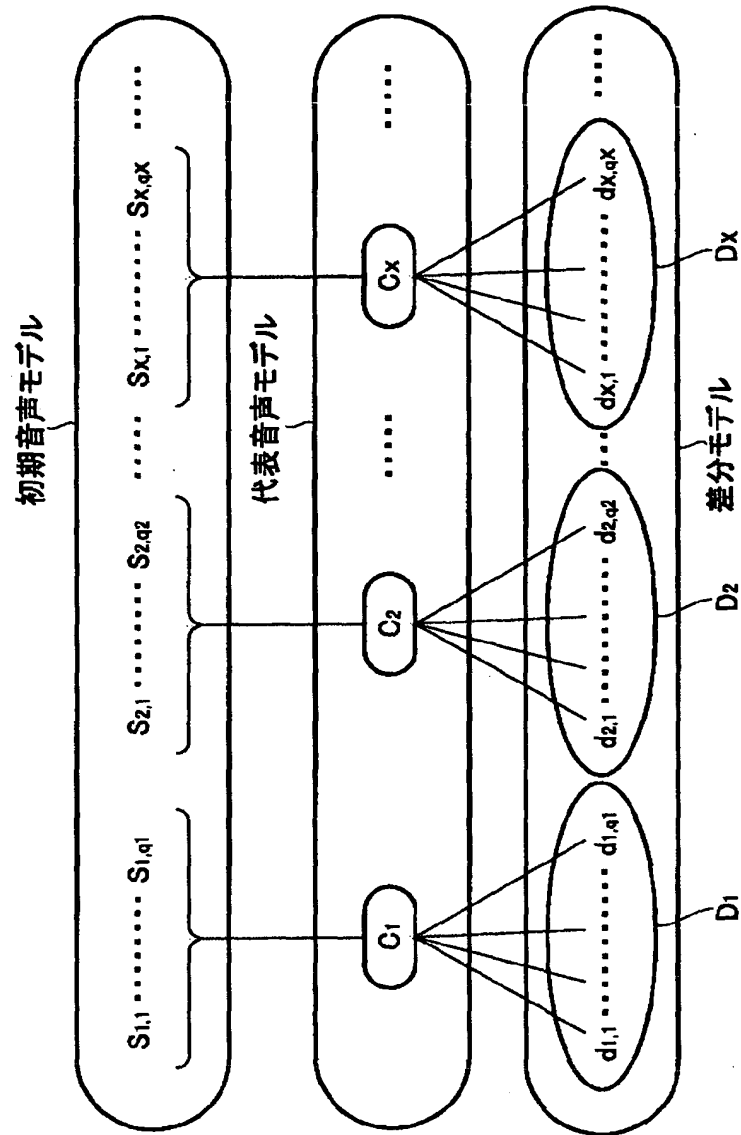
【図1】



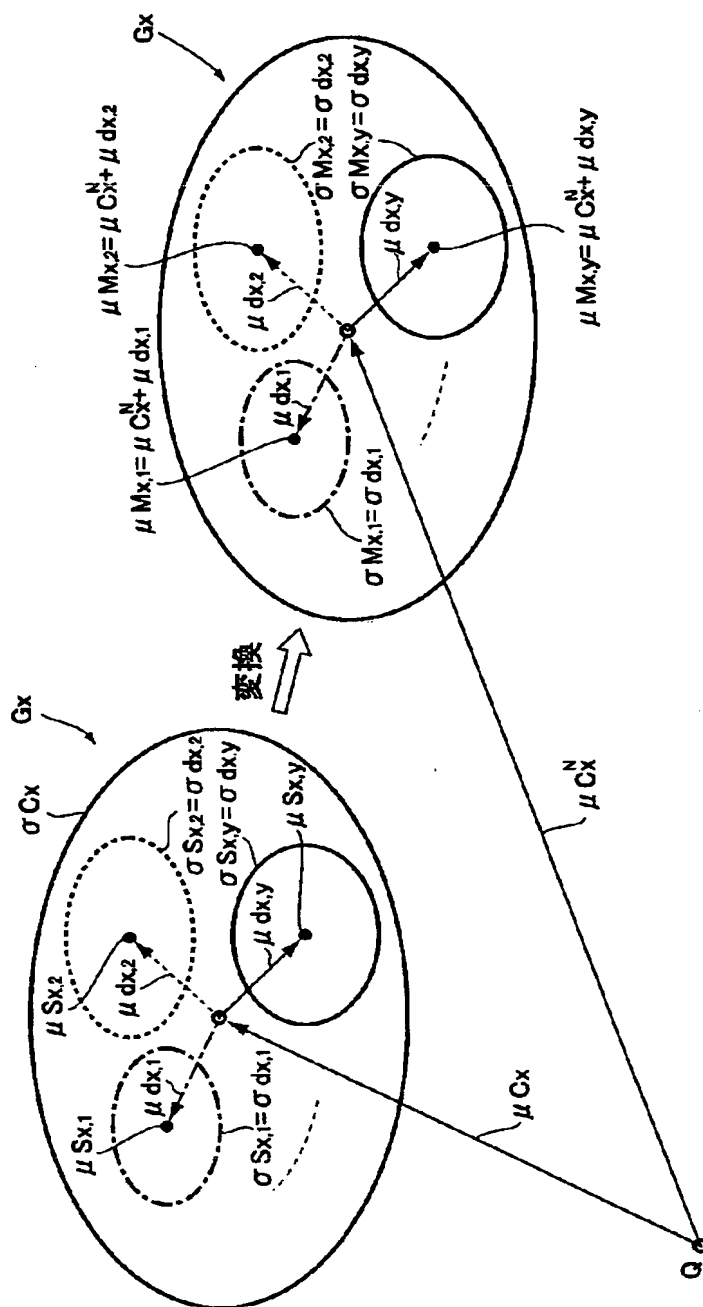
【図 2】



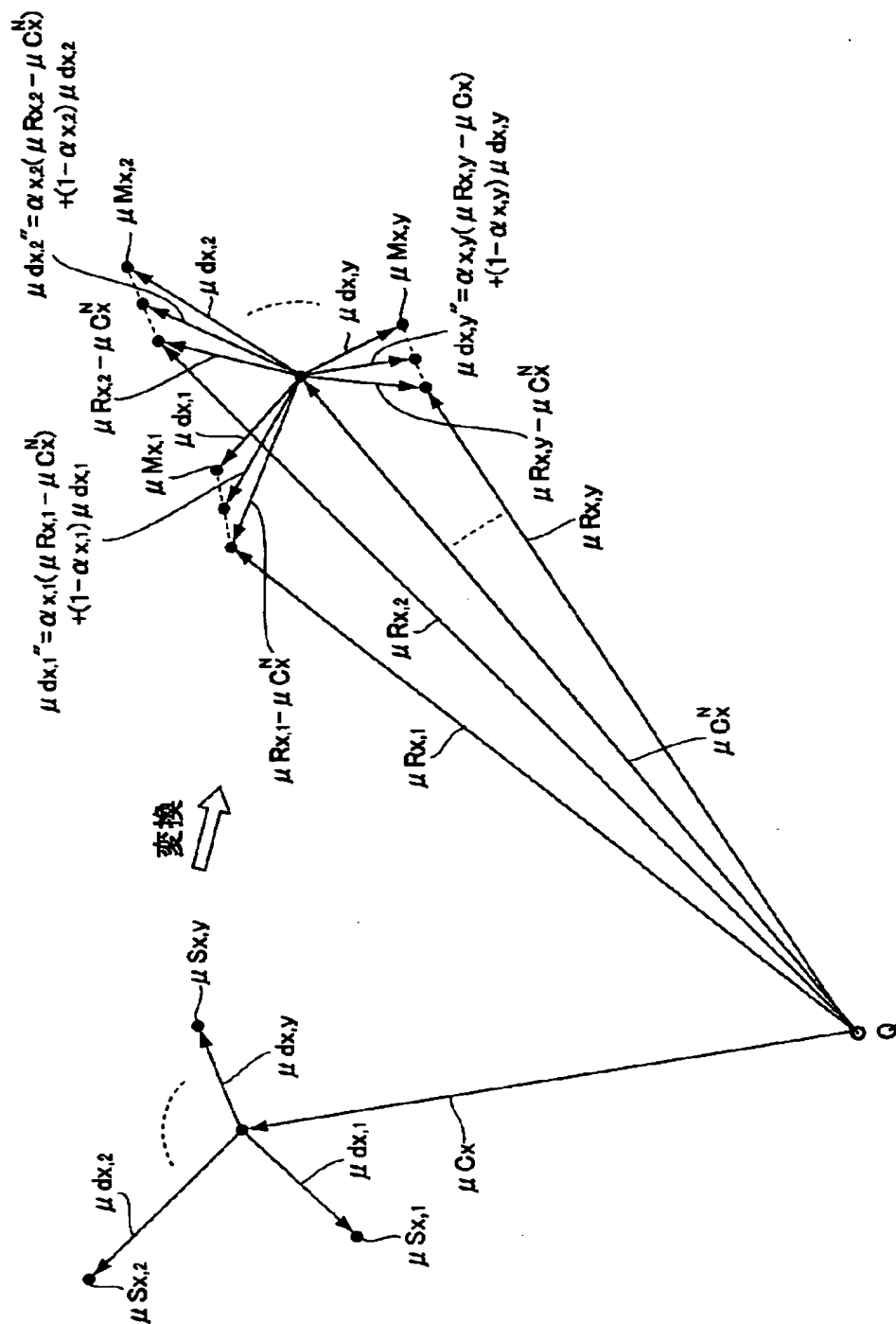
【図 3】



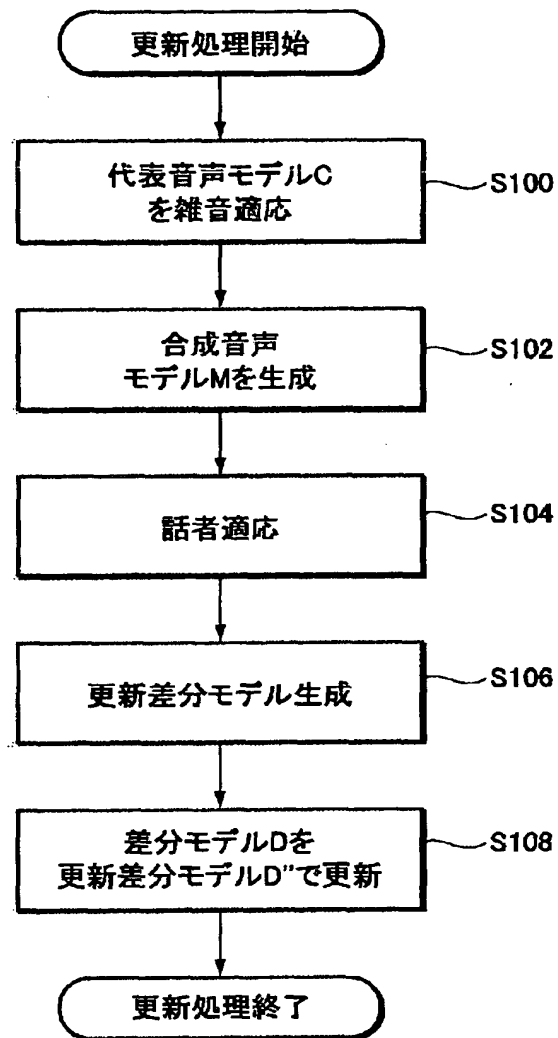
【図4】



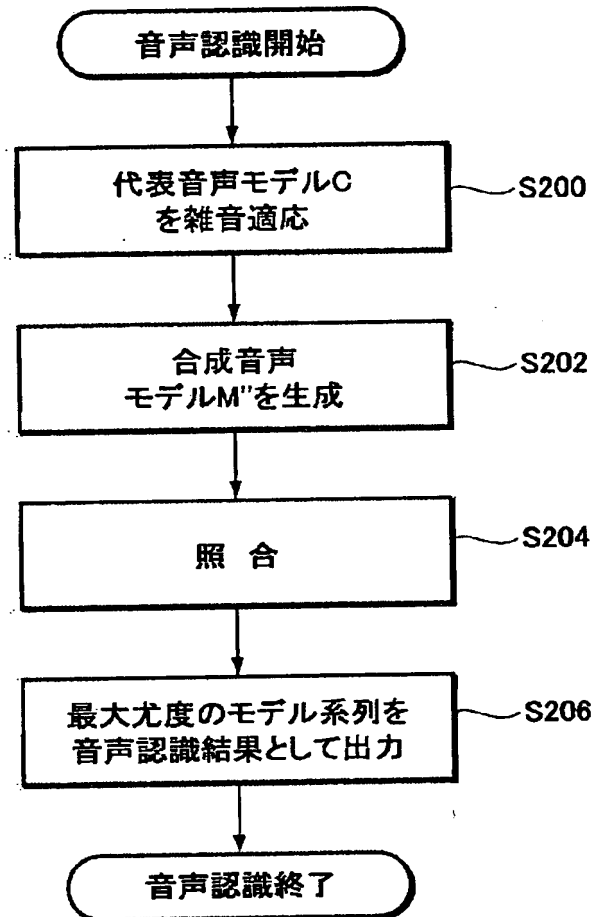
【図 5】



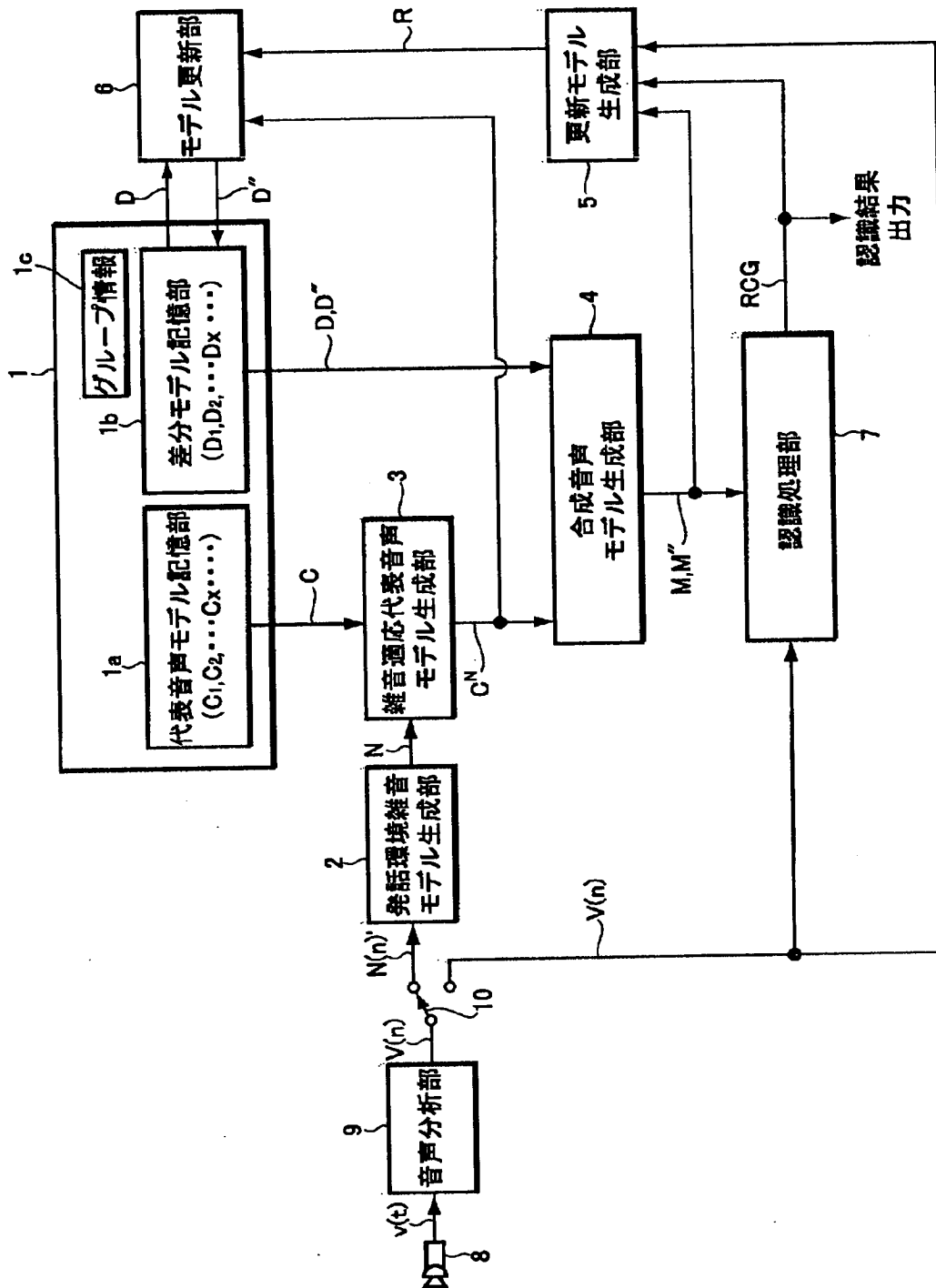
【図 6】



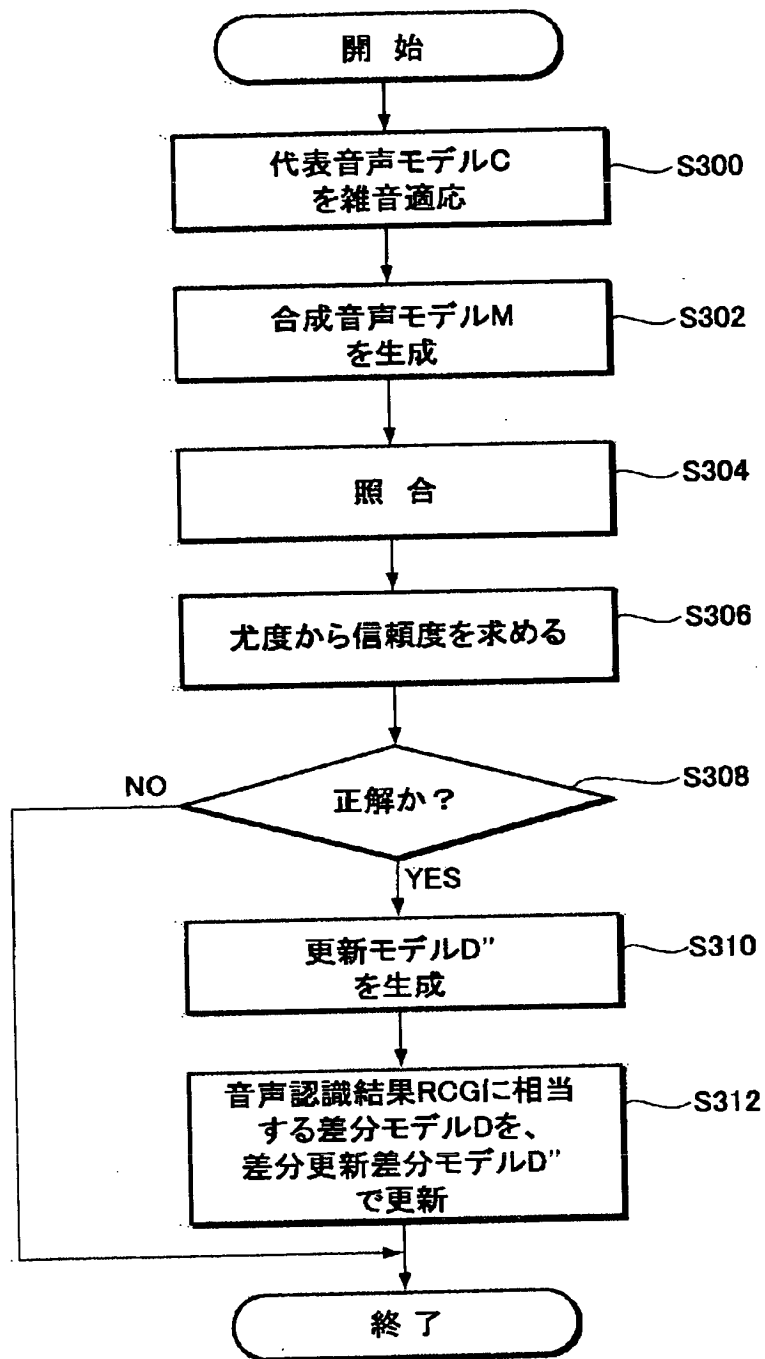
【図 7】



【図8】



【図9】



【書類名】 要約書

【要約】

【課題】 雑音適応と話者適応の処理量を低減する。

【解決手段】 初期音声モデルをクラスタリング等することで得られる代表音声モデルCと差分モデルDを予め代表音声モデル記憶部1aと差分モデル記憶部1bに記憶しておく。音声認識を行う前に、代表音声モデルCに対し雑音適応を施して雑音適応代表音声モデル C^N を生成し、差分モデルDを合成することで雑音適応を施した合成音声モデルMを生成する。その合成音声モデルMに対し発話音声の特徴ベクトル系列 $V(n)$ により話者適応を施し雑音話者適応音声モデルRを生成する。雑音話者適応音声モデルRと雑音適応代表音声モデル C^N との差分から更新差分モデル D'' を生成し、更新差分モデル D'' で記憶部1bの差分モデルDを更新する。音声認識に際して、代表音声モデルCと更新差分モデル D'' を合成することで雑音適応及び話者適応を施した合成音声モデル M'' を生成し、認識すべき話者音声の特徴ベクトル系列 $V(n)$ を照合して音声認識を行う。

【選択図】 図1

出 願 人 履 歴 情 報

識別番号 [000005016]

1. 変更年月日 1990年 8月31日
[変更理由] 新規登録
住 所 東京都目黒区目黒1丁目4番1号
氏 名 パイオニア株式会社